

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL



**Estudo da diversidade de vírus da imunodeficiência humana
tipo 1 em indivíduos infetados de naturalidade Guineense e
residentes em Portugal**

Sara Cebola Carrilho Basílio

Mestrado em Biologia Humana e Ambiente

Dissertação orientada por:
Inv^a Doutora Elizabeth Pádua
Instituto Nacional de Saúde, Dr. Ricardo Jorge
Prof^a Doutora Maria Teresa Rebelo
Faculdade de Ciências da Universidade de Lisboa

Agradecimentos

Gostaria de agradecer a todas as pessoas que, de algum modo, contribuíram para que pudesse realizar e concluir este mestrado:

- * Em primeiro lugar, ao Instituto Nacional de Saúde Dr. Ricardo Jorge pela oportunidade que dá aos alunos de realizarem os seus estudos na instituição;
- * À Doutora Elizabeth Pádua do INSA, por me ter proporcionado a oportunidade de realizar este estudo sob a sua supervisão, pelo apoio e conhecimentos transmitidos e sobretudo pela confiança, compreensão e paciência demonstrada ao longo destes meses;
- * À Professora Maria Teresa Rebelo, da Faculdade de Ciências da Universidade de Lisboa por ter aceitado a co-orientação deste projeto, pela sua disponibilidade, simpatia e pelos conselhos e sugestões;
- * À Catarina Almeida e Ivone Água-Doce, colaboradoras do Laboratório Nacional de Referência de VIH e Hepatites B e C, por toda a ajuda, disponibilidade e conhecimento laboratorial transmitidos e pela constante boa disposição com que me acolheram;
- * Aos elementos da Unidade de Tecnologia e Inovação (UTI) do INSA pelo apoio na sequenciação das amostras;
- * À colega Catarina, que foi uma grande ajuda e companhia durante estes meses;
- * A todos os meus amigos, em especial às melhores colegas de casa e ao meu namorado pela constante paciência, pela entreaajuda e motivação, por todos os momentos de convívio partilhados e principalmente pela amizade;
- * À minha família, principalmente à minha mãe e à minha avó, um agradecimento especial, pelo seu apoio incondicional em todos os aspetos, por todo o esforço e compreensão que sempre demonstraram;

A todos os mencionados e ainda a todos os que de algum modo marcaram a realização deste trabalho,

MUITO OBRIGADA!

Resumo

O vírus da imunodeficiência humana (VIH) é caracterizado por uma elevada diversidade genética que resulta principalmente das taxas de mutação e recombinação genética que origina novas de variantes genómicas ao longo do tempo. Este facto, juntamente com o aparecimento de mutações de resistência aos antirretrovirais tem causado diversos problemas ao nível do diagnóstico e tratamento da infeção.

O presente estudo pretendeu conhecer a diversidade genética do VIH-1 num grupo de 42 mulheres de naturalidade Guineense e residentes em Portugal e comparar os resultados com os obtidos num estudo anteriormente desenvolvido em indivíduos nativos e residentes na Guiné-Bissau.

Para caracterizar as amostras foram amplificadas por *nested*-PCR as regiões *env*, *nef*, PR e RT do VIH-1, posteriormente sequenciadas e realizada a análise filogenética das sequências com recurso a diferentes programas bioinformáticos. Foi realizada a análise das sequências aminoacídicas das quatro regiões genómicas, observando o grau de conservação e disrupção dos principais domínios estruturais e funcionais das proteínas, e ainda, a deteção das mutações de resistência associadas aos inibidores da PR e RT do VIH-1.

No grupo estudado foi obtida uma prevalência da CRF02_AG de 82% e identificados 5% de casos de infeção pelo sub-subtipo A1 (subtipo puro). Duas sequências foram classificadas em subtipo A, no entanto, limitado apenas à região *env* do VIH-1. Em três sequências foram observadas, diferentes recombinações (CRF02_AG/A1; CRF02_AG/J/A; 09_cpx/A) sendo classificadas como potencialmente formas recombinantes únicas.

Foi ainda observado, e termos gerais uma conservação dos domínios funcionais e estruturais das proteínas. Foi identificada elevada percentagem de polimorfismos naturais, associados a subtipos não-B. Foram detetadas mutações de resistência em 16% das sequências analisadas, associadas tanto aos inibidores da protease (V11I, T74P e N88T) como aos inibidores da transcriptase reversa (A98G, K103N e M184V).

Neste estudo, observou-se uma diversidade genética do VIH-1 semelhante à encontrada no estudo realizado em amostras de nativos e residentes na Guiné-Bissau e bastante inferior à diversidade descrita em Portugal. No entanto, esta diversidade de contribui para um padrão molecular da infeção no nosso país, único entre os países Europeus.

Palavras-Chave: VIH-1, Portugal, Guiné-Bissau, Diversidade genética, Regiões genómicas

Abstract

The human immunodeficiency virus (HIV) is characterized by a wide genetic diversity from higher rates of mutation and genetic recombination, originating over time, large number of genomic variants. This fact, together with the appearance of the virus with resistance mutations to the antiretroviral has caused many problems for the diagnosis, infection treatment and drug development.

The present study aims to determine the genetic diversity of HIV-1 in a group of 42 individuals of Guinean nationality and residents in Portugal, comparing the results with a study developed in the laboratory with native and resident individuals of Guinea -Bissau.

To characterize the samples, were amplified by nested-PCR the genomic regions, env, nef, PR and RT of the HIV-1, subsequently sequenced and held phylogenetic analysis of the sequences using different bioinformatic programs. Was performed an analyses of the amino acid sequences of the four regions of the study, observing the retention degree and disruption of the main structural and functional domains of HIV-1, and also the detection of the resistance mutations associated with the PR and RT inhibitors.

In the studied group was observed a prevalence of the CRF02_AG in 82% and identified 5% of cases of infection by the sub-subtype A1 (pure subtype). Two sequences were classified as subtype A, however only limited to the env region of HIV-1. In three sequences were observed different recombinations (CRF02_AG/A1; CRF02_AG/J/A; 09_cpx/A) being classified as potentially unique recombinant forms, having a mosaic genome.

It was also observed, in general, a conservation of functional and structural domains of proteins. It was identified high percentage of polymorphisms associated with non-B subtypes. Resistance mutations were detected in 16% of analysed sequences associated with both protease inhibitors (V11I, T74P and N88T) as the reverse transcriptase inhibitors (A98G, K103N and M184V).

In this study, we observed a genetic diversity of HIV-1 similar to that found in the study conducted on samples of native and resident Guinea-Bissau and much lower than the diversity described in Portugal. However, this diversity contributes to a molecular pattern of infection in our country, unique among European countries

Key words: VIH-1, Portugal, Guinea-Bissau, Genetic Diversity, Genomic Regions

Índice Geral

Agradecimentos	I
Resumo.....	II
Abstract	III
Índice Geral.....	IV
Índice de Figuras	VI
Índice de Tabelas.....	VII
Lista de Siglas e Abreviaturas.....	VIII
1. Introdução	1
1.1 A descoberta e a origem do Vírus da Imunodeficiência Humana	2
1.2 Classificação morfológica e organização genética do VIH-1	3
1.2.1 Ciclo replicativo	4
1.2.2 Transmissão e Progressão da infeção	5
1.2.3 Diagnóstico e Tratamento	7
1.3 Diversidade genética do VIH-1	8
1.3.1 Classificação dos vírus e sua distribuição global	8
1.3.2 Causas e consequências da variabilidade genética	10
1.4 Perfil molecular da infeção VIH	10
1.4.1 Na Guiné Bissau	10
1.4.2 Em Portugal	12
1.5 Genes <i>env</i> , <i>nef</i> e <i>pol</i> do VIH-1	13
1.5.1 Gene <i>env</i> e glicoproteína <i>gp120</i> (SU)	13
1.5.2 Gene <i>nef</i> e Proteína Nef	15
1.5.3 Gene <i>pol</i> e as Proteínas PR e RT	16
1.5.3.1 Protease	16
1.5.3.2 Transcriptase Reversa	17
1.6 Inibidores da Protease e Transcriptase Reversa	18
2. Enquadramento e Objetivos do Estudo	20
3. Material e Métodos.....	23
3.1 População estudada	24
3.2 Extração de ADN proviral do VIH-1	24
3.3 Caracterização dos <i>primers</i> utilizados no estudo.....	24
3.3.1 Primers para a amplificação das regiões <i>env</i> , <i>nef</i> , PR e RT do VIH-1	26

3.4 Condições das reações de amplificação do VIH-1	26
3.5 Visualização de produtos amplificados (2ª amplificação)	28
3.6 Purificação e Sequenciação dos produtos de PCR	29
3.6.1 Purificação com o <i>kit</i> ExoSap- IT USB	29
3.6.2 Purificação com o <i>kit</i> QIAquick® PCR Purification	29
3.6.3 Sequenciação dos purificados com o <i>kit</i> Big Dye™ Terminator Cycle Sequencing Ready Reaction.....	30
3.7 Análise de Sequências Nucleotídicas.....	31
3.7.1 Cromatogramas e Construção de Sequências Consenso	31
3.7.2 Inferência Filogenética	31
3.7.3. Distâncias Nucleotídicas.....	31
3.8 Sequências aminoacídicas de V3-V5, proteína Nef, PR e RT	32
4. Resultados	33
4.1 Caracterização das Amostras Estudadas	34
4.2 Amplificação das regiões env, nef, PR e RT de VIH-1	34
4.3 Caraterização Molecular das Sequências de VIH-1	35
4.3.1 Análise Filogenética da Região V3-V5 de <i>env</i>	35
4.3.2 Análise Filogenética da Região nef.....	38
4.3.3 Análise Filogenética da Região PR.....	40
4.3.4 Análise Filogenética da Região RT.....	44
4.3.5 Análise Conjunta das Regiões V3-V5, nef, PR e RT.....	45
4.3.6. Distâncias nucleotídicas entre sequências da região V3-V5 de env e gene nef	47
4.4 Análise das Sequências Aminoacídicas de VIH-1.....	48
4.4.1 Região V3-V5	48
4.4.2. Proteína Nef	50
4.4.3. Região PR.....	52
4.4.4. Região RT.....	55
5. Discussão	56
6. Conclusão	66
7. Referências	69
8. Anexos	79

Índice de Figuras

1. Introdução

Figura 1.1. Relações filogenéticas entre vírus que infeta primatas não humanos e humanos	3
Figura 1.2. Representação esquemática da partícula de VIH e das suas principais proteínas.....	4
Figura 1.3. Representação esquemática dos principais processos da replicação do VIH.	5
Figura 1.4. Frequência de Subtipos e CRF do VIH-1 existente na base de dados em circulação na população da Guiné-Bissau.....	12
Figura 1.5. Representação esquemática dos domínios internos e externos da gp120 de VIH-1.	14
Figura 1.6. Principais motivos estruturais e funcionais da proteína Nef do VIH-1.	16
Figura 1.7. RX da análise cristalográfica do centro ativo da região da protease do VIH-1.	17

4. Resultados

Figura 4.1. Visualização dos produtos de PCR	35
Figura 4.2. Árvore filogenética da região V3-V5 do VIH-1.....	36/37
Figura 4.3. Árvore filogenética para a região <i>nef</i> do VIH-1.	39/40
Figura 4.4. Árvore filogenética da região PR do gene <i>pol</i> do VIH-1.	42
Figura 4.5. Árvore filogenética da região RT do gene <i>pol</i> do VIH-1	44
Figura 4.6. Alinhamento parcial da região V3-V5 de <i>env</i>	48
Figura 4.7. Alinhamento parcial de Nef.	52
Figura 4.8. Alinhamento da Região PR.....	53
Figura 4.9. Alinhamento parcial da região RT	54

Índice de Tabelas

3. Material e Métodos

Tabela 3.1. Primers utilizados nas misturas de reação para amplificação das diferentes regiões genômicas do VIH-1	25
Tabela 3.2. Programas de amplificação (1ª e 2ª reação de PCR) utilizados para a amplificação da região <i>env</i> do VIH-1	27
Tabela 3.3. Programa de amplificação (1ª e 2ª reação de PCR) utilizado na amplificação do gene de <i>nef</i> do HIV-1	28
Tabela 3.4. Programa de amplificação (1ª e 2ª reação de PCR) utilizado na amplificação de PR e de RT do VIH-1	28
Tabela 3.5- Programa de sequenciação nucleotídica.	30

4. Resultados

Tabela 4.1. Classificação filogenética das sequências GP em estudo.....	46
Tabela 4.2. Cálculo das distâncias nucleotídicas intra e inter grupo para as sequências V3-V5 e <i>nef</i> do VIH-1.....	47
Tabela 4.3. Polimorfismos encontrados na sequência aminoacídica V3 do VIH-1.	49
Tabela 4.4. Previsão do co-recetor utilizado nas amostras GP em estudo	50
Tabela 4.5. Polimorfismos encontrados na sequência aminoacídica de Nef do VIH-1	51
Tabela 4.6. Frequência das mutações de resistência aos PI.....	53
Tabela 4.7. Frequência das mutações de resistência aos NNRTI e NRTI	55

Lista de Siglas e Abreviaturas

- 3TC** - Lamiduvina
ABC - Abacavir
ADN - Ácido desoxirribonucleico
ARN - Ácido ribonucleico
ARNm - Ácido ribonucleico mensageiro
ATV - Atazanavir
BLAST – do inglês, *Basic Local Alignment and Search Tool*
CCR5 – Recetor C-C da quimiocina tipo 5
CD4 - Recetor celular, do inglês *Cluster of differentiation type 4*
CMSP – Células Mononucleares de Sangue Periférico
CRF - Formas recombinantes em circulação, do inglês *Circulating recombinant form*
CTL – Linfócitos Citolíticos
CXCR4 –C recetor C-X-da quimiocina tipo 4
dNTP – Desoxirribonucleotídeos fosfatados
DRV – Darunavir
FTC - Emtricitabine
Gp – Glicoproteína
HAART - Terapêutica antirretroviral de alta eficácia, do inglês *Highly active antiretroviral therapy*
IN - Integrase
INSA - Instituto Nacional de Saúde Dr. Ricardo Jorge
Kb – Quilo Bases
MA – Matriz Proteica
MEGA - do inglês, *Molecular Evolutionary Genetic Analysis*
MHC-I - Complexo Principal de Histocompatibilidade de Classe I, do inglês *Major Histocompatibility Complex Class I*
MHC-II - Complexo Principal de Histocompatibilidade de Classe II, do inglês *Major Histocompatibility Complex Class II*
mM – milimolar
NC – Proteína da Nucleocápside
nt- Nucleótidos
NFV – Nelfinavir
NNRTI - Inibidores não nucleosídicos da transcriptase reversa, do inglês *Non-nucleoside reverse transcriptase inhibitors*
NRTI - Inibidores nucleosídicos da transcriptase reversa, do inglês *Nucleoside reverse transcriptase inhibitors*
NSI - Variantes virais não indutoras de sincícios, do inglês *Non-Syncytium Inducing*

Pb – Pares de Bases

PCR - Reação em cadeia da polimerase de ADN, do inglês *Polymerase Chain Reaction*

PI – Inibidor da Protease, do inglês *Protease inhibitors*

PR - Protease

RNase H - Atividade da Ribonuclease H

RT – Transcriptase Reversa

SI - Variantes virais indutoras de sincícios, do inglês *Syncytium Inducing*

SIDA - Síndrome da Imunodeficiência Adquirida

SU – Glicoproteína de Superfície

TM - Glicoproteína Transmembranar

UDI – Utilizadores de Drogas Injetáveis

URF – Formas recombinantes únicas, do inglês, *Unique recombinant forms*

V – Volts

VIH-1 - Vírus da Imunodeficiência Humana tipo 1

VIH-2 - Vírus da Imunodeficiência Humana tipo 2

VIS_{cpz} - Vírus da Imunodeficiência Símia de chimpanzés

VIS_{sm} - Vírus da Imunodeficiência Símia de *Sooty mangabeys*

μl - Microlitro

μM - Micromolar

1. Introdução

1.1 A descoberta e a origem do Vírus da Imunodeficiência Humana

O vírus da imunodeficiência humana (VIH) é o agente etiológico da Síndrome da Imunodeficiência Adquirida (SIDA). A infecção por este vírus continua a ser um dos maiores problemas de saúde pública no mundo, sendo a infecção caracterizada por elevadas taxas de morbidade e mortalidade na população humana, particularmente observadas no continente africano, onde existe ainda um acesso limitado ao tratamento (Friedrich *et al*, 2011). De acordo com dados da Organização Mundial de Saúde, existem cerca de 36,9 milhões de pessoas infetadas no mundo. Estima-se que desde o início da epidemia VIH/SIDA mais de 71 milhões de pessoas se infetaram e cerca de 34 milhões não resistiram à virulência da infecção (<http://www.who.int/gho/hiv/en/>).

A descoberta do primeiro vírus causador da SIDA ocorreu em 1983, quando o vírus da imunodeficiência humana tipo 1 (VIH-1) foi isolado de um doente com SIDA pela equipa de Luc Montagnier do Instituto Pasteur de Paris (Barré-Sinoussi *et al*, 1983). Em 1986, a mesma equipa francesa em colaboração com clínicos e investigadores portugueses isolaram o segundo tipo de vírus causador da SIDA, posteriormente denominado de vírus da imunodeficiência tipo 2 (VIH-2) (Clavel *et al*, 1986). Apesar de ambos os vírus apresentarem uma homologia genómica de 40 a 60% entre si e induzirem respostas imunitárias cruzadas, observam-se diferenças significativas nas suas características antigénicas, biológicas e moleculares (Stebbing & Moyle, 2003; Antunes & Doroana, 2012).

A origem provável da infecção VIH no Homem deve-se à transmissão zoonótica a partir de primatas não humanos infetados com o vírus da imunodeficiência símia (VIS) (Hemelaar, 2012). Estes primatas constituem um reservatório do vírus com elevada potencialidade de infetar outras espécies animais, incluindo a espécie humana (Requejo, 2006). Estudos de filogenia revelaram um agrupamento robusto entre vírus da imunodeficiência humana e os vírus da imunodeficiência símia que infetam primatas não humanos, constituindo um ramo filogenético com uma origem comum, mostrando uma adaptação e evolução do vírus ao novo hospedeiro (Krogstad, 2003; Sharp & Hahn, 2011). A análise de sequências moleculares (genes ou genoma total do vírus) revelou que o VIH-1 se encontra próximo e filogeneticamente relacionado com VIScpz que infeta chimpanzés da subespécie *Pan troglodytes troglodytes*, (Gao *et al*, 1999) e o VIH-2 apresenta um elevado grau de homologia com VISsm que infeta

macacos da espécie *Cercocebus atys*, endêmica na África ocidental (Butler *et al*, 2007) (Figura 1.1).

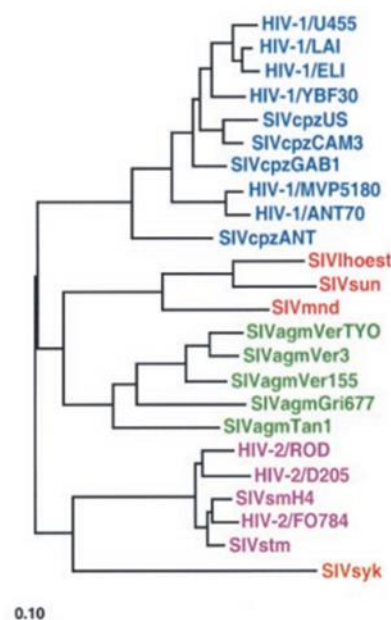


Figura 1.1. Análise de relações filogenéticas entre vírus que infeta primatas não humanos (SIV) e humanos (VIH). O agrupamento azul mostra a homologia existente entre o VIH-1 e o VIScpz e o agrupamento rosa entre o VIH-2 e o VISsm. Adaptado de Hahn *et al*, 2000.

A transmissão do vírus inter-espécie, a partir do macaco para o Homem pode ser fundamentada pelos riscos de infecção decorrentes de acidentes na caça frequente destes animais para alimentação humana em muitos países Africanos, assim como o seu cativeiro como animais de estimação (Gao *et al*, 1999).

1.2 Classificação morfológica e organização genética do VIH-1

O comité internacional para a taxonomia dos vírus classificou o VIH como pertencendo à família *Retroviridae*, sub-família *Orthoretrovirinae* e género *Lentivirus* (http://www.ictvonline.org/virusTaxonomy.asp?taxnode_id=20040149).

As partículas de VIH-1 têm uma morfologia esférica com cerca de 100 nm de diâmetro (Sierra *et al*, 2005). Todos os lentivírus possuem um invólucro constituído por uma bicamada lipídica que deriva da membrana da célula hospedeira. No invólucro encontram-se ancoradas importantes glicoproteínas, nomeadamente, a glicoproteína de superfície (gp120, SU) e a glicoproteína transmembranar (gp41, TM), bem como outras proteínas de origem celular

(Krogstad, 2003). O invólucro envolve a matriz proteica (p17, MA), essencial para a integridade do virião, que por sua vez reveste cápside viral com a forma de um cone truncado. No seu interior, para além das duas moléculas genómicas de ácido ribonucleico (ARN), estabilizado pela proteína da nucleocápside (p7, NC), possui enzimas essenciais para o ciclo replicativo tais como, a protease (PR), a transcriptase reversa (RT) e a integrase (IN), e ainda, as proteínas acessórias como Nef, Vif e Vpr (Figura 1.2) (Turner & Summers, 1999; Sierra *et al*, 2005).

O VIH possui um genoma de cerca de 9,2 kb constituído por duas moléculas de ARN lineares, de cadeia simples e com polaridade positiva (Krogstad, 2003) que codifica as principais proteínas estruturais e não estruturais que são comuns em todos os retrovírus. Da

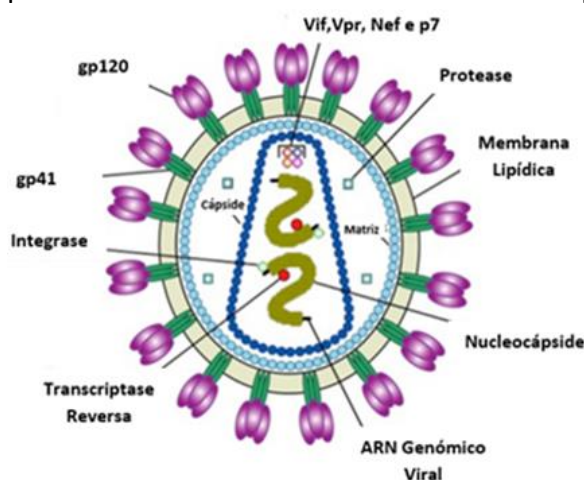


Figura 1.2. Representação esquemática da partícula de VIH e das suas principais proteínas. Adaptado de http://www.ff.ul.pt/~jazevedo/Site/Mest_CBF_files/HIV-JMAP.pdf

extremidade 5' para a extremidade 3' do genoma do vírus são identificados os genes principais, *gag*, *pol* e *env* que codificam grande parte dos componentes estruturais, enzimáticos e do invólucro do virião (Freed, 2001). Para além destes genes são identificados no genoma do VIH-1 os genes reguladores *tat* e *rev* assim como e os genes acessórios (*vpr*, *vif*, *vpu*, e *nef*) (Malin & Emerman, 2008; Antunes & Doroana, 2012).

1.2.1 Ciclo replicativo

O ciclo de replicação do VIH compreende várias etapas que se podem agrupar em duas fases distintas. Uma primeira fase que corresponde à infeção com entrada do vírus na célula e a uma segunda fase que corresponde à expressão génica e a consequente formação de novos viriões (Antunes e Doroana, 2012). A fase de infeção inicia-se com a ligação do virião à molécula de CD4 presente na superfície da célula-alvo (linfócitos T auxiliares, monócitos ou macrófagos) e a um coreceptor (geralmente o recetor das quimiocinas, CCR5 ou CXCR4) (Berger *et al*, 1999; Engelman, 2012). Como resultado desta ligação, ocorre a fusão do invólucro viral e a membrana celular com a entrada e degradação da cápside no citoplasma e consequente libertação do ARN viral que é convertido pela RT em ácido desoxirribonucleico

(ADN) de cadeia dupla (Freed, 2001; Antunes e Doroana, 2012). Num processo desencadeado pela integrase, o ADN é integrado no genoma da célula sendo posteriormente iniciada a transcrição do provírus, marcando o início da segunda fase do ciclo replicativo do VIH (Freed, 2001). A transcrição do provírus conduz à síntese de moléculas de ARN mensageiro (ARNm), quer completas para a formação do genoma do vírus, quer traduzidas nos precursores das várias proteínas virais (Sierra *et al*, 2005; Antunes & Doroana, 2012). O ARN genómico e as proteínas virais são transportadas e montadas junto da membrana celular. Os viriões adquirem invólucro por um processo de gemulação sendo libertados da célula ainda na forma imatura. Já fora da célula, por ação da protease, ocorre a maturação da partícula, processo essencial para tornar os viriões infecciosos e capazes de infetar novas células, completando-se deste modo, o ciclo replicativo do VIH (Figura 1.3) (Antunes & Doroana, 2012).

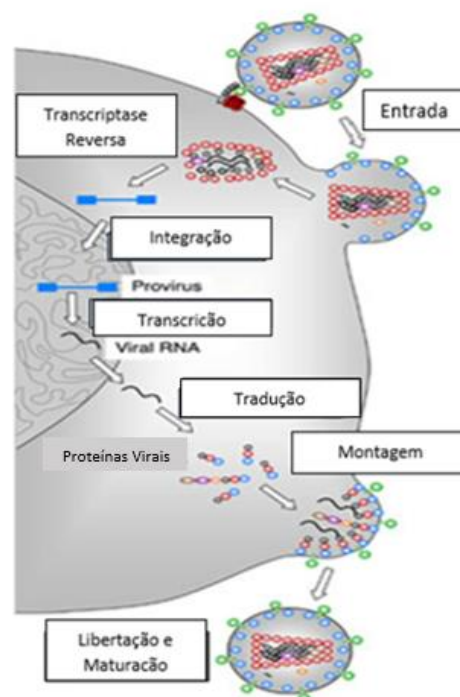


Figura 1.3. Representação esquemática dos principais processos da replicação do VIH. Adaptado de http://www.adarc.org/hiv_replication_632.html

1.2.2 Transmissão e Progressão da infeção

A transmissão do VIH-1 depende de vários fatores, tais como, as taxas de replicação e infecciosidade do vírus, a sua concentração nos fluídos corporais e o tempo de exposição do indivíduo. A natureza e os níveis de suscetibilidade celular, as condições imunológicas e clínicas do hospedeiro são também importantes na infeção e transmissão do VIH-1. Outros fatores podem influenciar o decurso da infeção e a consequente progressão para doença, nomeadamente as infeções oportunistas e o estado nutricional dos indivíduos (Levy, 2009; Frieden *et al*, 2015).

As principais formas de transmissão do VIH-1 são a via sexual e a via sanguínea. A partilha de seringas ou material de injeção contaminado usado no consumo de drogas ilícitas por via endovenosa tem sido uma das maiores fontes de aquisição do VIH. Na ausência de medidas

de prevenção, a transmissão mãe-filho do VIH-1 é frequente, podendo ocorrer a transmissão do vírus ao feto durante a gravidez, durante o trabalho de parto, ou ainda, após o parto através da amamentação da criança com leite materno (Cohen *et al*, 2008; Frieden *et al*, 2015).

A infeção natural por VIH-1 apresenta fases de evolução clinicamente distintas: a fase aguda, a fase crónica e a fase terminal de SIDA. A progressão para doença apresenta uma grande variabilidade entre os indivíduos infetados. Alguns podem progredir rapidamente (entre 1 a 3 anos) da fase aguda para SIDA, conhecidos como “*fast progressors*”, enquanto outros podem permanecer assintomáticos durante mais de 10 anos, sem evidência significativa de disfunção imune, sendo denominados de “*long-term non-progressors*” (Singh *et al*, 2008). Existem ainda casos raros descritos na literatura, de indivíduos denominados “*elite controllers*” ou “*HIV controllers*”, que mesmo na ausência de tratamento possuem a capacidade de controlar a replicação do vírus para níveis abaixo do limiar de deteção das técnicas usadas na monitorização da infeção, mantendo-se imunologicamente estáveis (Pereyra *et al*, 2009).

A fase aguda inicia-se imediatamente após a infeção do indivíduo, observando-se o aumento da carga viral na circulação plasmática e o reconhecimento do VIH pelo sistema imunitário e a diminuição das células T CD4+. Estas células são suscetíveis à infeção tornando-se o alvo principal do VIH (Levy, 2009). Esta fase é assintomática em cerca de 50% dos infetados e possui uma duração de cerca de 4 a 6 semanas. Os restantes indivíduos podem desenvolver sintomas semelhantes a uma gripe ou uma mononucleose infecciosa. Na fase crónica da infeção, a resposta imunitária celular e a síntese de anticorpos contra o VIH conduzem a uma redução da carga viral no plasma, sendo esta fase caracterizada por uma latência clínica (Berger *et al*, 1999). No entanto, durante este período existe uma replicação persistente do VIH nos gânglios linfáticos e uma contínua degradação das condições imunológicas no indivíduo infetado (Antunes & Doroana, 2012). A fase de SIDA é caracterizada essencialmente pelo aumento exponencial da replicação viral e por uma diminuição abaixo de 200 células/ μ L dos níveis de linfócitos T CD4+ (Sierra *et al*, 2005; Levy, 2009). O enfraquecimento do sistema imunitário favorece o aparecimento de infeções oportunistas, que nos indivíduos mais debilitados pode conduzir à morte.

1.2.3 Diagnóstico e Tratamento

O diagnóstico da infecção VIH-1 é maioritariamente realizado pela deteção de anticorpos específicos contra o vírus a partir de plasma ou soro humano. Em determinadas condições, tais como em infeções adquiridas muito recentemente (2 a 4 semanas) ou nos recém-nascidos (<12 meses de idade) de mães infetadas, o diagnóstico deve ser realizado pela deteção direta do vírus por reações de PCR com amplificação dos ácidos nucleicos do VIH (Fearon, 2005; Simon *et al*, 2006).

O diagnóstico precoce da infecção VIH conduz a benefícios para a saúde individual e pública, sendo mundialmente reconhecido que o tratamento dos infetados pode constituir também uma estratégia de prevenção diminuindo o risco de disseminação de VIH na comunidade (Fearon, 2005; Cornett & Kirn, 2013).

No entanto, o VIH é caracterizado por apresentar uma elevada diversidade genética dificultando o diagnóstico laboratorial da infeção. Ao longo dos anos tem sido descrito na literatura a ocorrência de resultados falso-negativos devido ao fraco desempenho de alguns testes de diagnóstico para a deteção de diferentes grupos, subtipos ou formas recombinantes em circulação de VIH-1 (Butler *et al*, 2007).

Na década de 80, o aparecimento dos primeiros antirretrovirais que pareciam controlar a replicação viral, levou a crer que o tratamento poderia conduzir a uma potencial cura do VIH (Castro-Nallar *et al*, 2012). Contudo, o aparecimento continuado de novas variantes do VIH, algumas com reduzida suscetibilidade ou resistência aos antirretrovirais, acabou por se tornar um dos maiores obstáculos à eficácia do tratamento a longo prazo (Kandathil *et al*, 2005). Este facto é ainda considerado um grave problema devido à complexa gestão médica e terapêutica dos indivíduos infetados, particularmente, por permitir também uma contínua transmissibilidade de variantes resistentes na população (Thomson *et al*, 2002).

Atualmente, os regimes terapêuticos compreendem uma combinação de antirretrovirais de diferentes classes, que atuam em diferentes fases do ciclo replicativo do VIH e conduzem a maior eficácia do tratamento. Porém, ainda é essencial a continuação do desenvolvimento de

novos fármacos de maior potência que possam alcançar e destruir o vírus nos reservatórios e assim permitir a eliminação do vírus do organismo (Menéndez-Arias, 2010).

1.3 Diversidade genética do VIH-1

O VIH-1 é caracterizado por uma elevada diversidade genética e rápida evolução no tempo. É reconhecido que a recombinação genética entre subtipos do VIH-1 contribui significativamente para a evasão do vírus ao sistema imunológico do hospedeiro, e consequentemente, a sua transmissão e disseminação na população humana (Osmanov *et al*, 2002).

1.3.1 Classificação dos vírus e sua distribuição global

O VIH-1 encontra-se dividido em quatro grupos: o grupo M, o grupo N, o grupo O e o grupo P (Plantier *et al*, 2009).

Contrariamente aos grupos N, O e P que representam infeções esporádicas maioritariamente restritas à África Central (Andersson *et al*, 1999), o grupo M representa a pandemia da infeção e inclui os vírus responsáveis por cerca de 90% das infeções por VIH-1 no mundo (Gupta *et al*, 2005). O grupo M é composto por nove subtipos virais, denominados A, B, C, D, F, G, H, J e K, encontrando-se os subtipos A e F subdivididos em sub-subtipos, respetivamente de A1 a A4 e de F1 a F2 (Pinto & Struchiner, 2006; Tebit & Arts, 2011).

A sequenciação do genoma de VIH-1 permitiu também a identificação de formas recombinantes únicas (URF) e formas recombinantes em circulação (CRF), ambas resultam da recombinação do genoma entre vírus de subtipos diferentes, ou entre subtipos e recombinantes (Taylor *et al*, 2008). As URF constituem uma variante genómica identificada num único indivíduo, não se conhecendo a sua eventual capacidade de disseminação na população (forma única que não está em circulação). As CRF são variantes identificadas em dois ou mais indivíduos, epidemiologicamente não relacionados, estimando-se deste modo, o seu estabelecimento e transmissibilidade na população (forma em circulação). Cada vez mais é reconhecida a importância das CRF na medida que representam uma elevada incidência de infeção VIH-1 (Thomson *et al*, 2002). Até ao momento são conhecidas mais de 70 CRF (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

O VIH apresenta um padrão molecular na distribuição geográfica no mundo (Freitas *et al*, 2013).

O subtipo C é o responsável pela maioria das infeções identificadas em todo o mundo, sendo prevalente no continente africano, particularmente em países como a África do Sul, Etiópia e Índia. As CRFs resultantes da recombinação entre este subtipo e o subtipo B, as CRF07_BC e CRF08_BC, são frequentemente encontradas na China, em utilizadores de drogas injetáveis (UDI). Recombinantes com pequenas frações no genoma, originárias do subtipo C são muito comuns na Tanzânia (Thomson *et al*, 2002; Buttler *et al*, 2007; Tebit & Arts, 2011).

O subtipo A representa maioritariamente as infeções em países do Centro, Este e Ocidente do continente Africano, como o Quénia e Uganda, e também em antigos países da União Soviética associado fortemente a infeções em UDI (Thomson *et al*, 2002). As CRFs mais representativas nestas regiões são a CRF01_AE e CRF02_AG (Stebbing & Moyle, 2003; Tebit & Arts, 2011). A primeira, a CRF01_AE tem uma ampla distribuição na África Central e no Sudoeste da Ásia (Stebbing & Moyle, 2003; Gupta *et al*, 2005; Maartens *et al*, 2014) e a segunda, a CRF02_AG é a mais frequentemente encontrada a nível mundial.

As infeções pelo subtipo B apesar de apresentarem uma baixa frequência em termos gerais (menor número de pessoas mundialmente infetadas) são predominantes na Europa, na América do Norte e na Austrália. Em vários países do sudeste asiático, como o Japão e Médio Oriente, e também na Rússia, as infeções por este subtipo estão associadas a grupos populacionais de homossexuais (Stebbing & Moyle, 2003; Maartens *et al*, 2014).

As infeções pelos restantes subtipos de VIH-1 têm uma menor prevalência, podendo ser identificadas em pequena escala e em locais específicos. O subtipo D distribui-se maioritariamente pela África Oriental. O subtipo F é predominante na Roménia, infetando crianças institucionalizadas que foram infetadas por procedimentos técnicos e clínicos inseguros, nomeadamente a reutilização de agulhas e seringa, durante a vacinação. Existe ainda uma pequena percentagem de infeções pelo subtipo F e recombinantes BF em circulação, no Brasil. O subtipo G representa uma parte significativa das infeções na África Ocidental e Central, existindo uma elevada prevalência deste subtipo em Portugal (Kandathil *et al*, 2005; Buttler *et al*, 2007; Carvalho *et al*, 2015).

1.3.2 Causas e consequências da variabilidade genética

A elevada variabilidade genética é uma característica do VIH-1 e resulta da acumulação de vários fatores, nomeadamente:

- 1) A elevada taxa de mutação; em consequência da elevada frequência de erros cometidos pela RT durante a síntese de ADN e que não são corrigidos devido à ausência da atividade de correção exonucleásica 3'-5' da enzima (Simonsen *et al*, 1998; Gupta *et al*, 2005; Abecasis *et al*, 2009; Smyth *et al*, 2012).
- 2) Um ciclo replicativo curto com produção de um elevado número de descendentes; estima-se que, diariamente, sejam produzidos e libertados na corrente sanguínea cerca de 10^{10} viriões infecciosos (Pinto & Struchiner, 2006; Taylor *et al*, 2008).
- 3) A elevada taxa de recombinação genética; o genoma dimérico do VIH-1 (duas cadeias simples de ARN) permite que a RT alterne entre as cadeias durante a síntese de ADN e origine descendentes com informação genómica mista e diferente da dos seus progenitores (Gupta *et al*, 2005, Abecasis *et al*, 2009).
- 4) A seleção, adaptação e evolução dos vírus; a pressão seletiva exercida nas populações virais, quer pelo sistema imunitário quer pela terapêutica antirretroviral, resultam na seleção de variantes e numa rápida taxa de evolução do vírus; a adaptação do vírus permite a manutenção da infeção persistente e a sua transmissão a novos hospedeiros (Abecasis *et al*, 2009; Eberle & Gurtler, 2012).

1.4 Perfil molecular da infeção VIH

1.4.1 Na Guiné Bissau

No início da epidemia foram descritos raros casos de infeção por VIH-1 na Guiné Bissau, sendo este país considerado um dos países da África Ocidental com maior prevalência e incidência da infeção por VIH-2 (Jespersen *et al*, 2015). No entanto, os estudos epidemiológicos realizados ao longo do tempo têm demonstrado que a prevalência do VIH-2 na Guiné Bissau decresceu a partir dos anos 80 e que durante um período de dez anos, passou de 10% para 5,9% (Larsen *et al*, 1998). A redução na proporção de casos de infeção VIH-2 e o aumento da prevalência do VIH-1 confirmada por estudos mais recentes é justificada por uma maior virulência e transmissibilidade do VIH-1 comparativamente ao VIH-2 (Silva *et al*, 2008).

Um dos primeiros estudos de epidemia molecular de VIH-1 na Guiné-Bissau mostrou que a maioria das sequências de VIH-1 estava estreitamente relacionada com os isolados IbNG, DJ263 e DJ264, partilhando um padrão de recombinação A/G comum (Andersson *et al*, 1999). Neste estudo foi descrito que esta variante é característica dos países da África Ocidental e que é diferente das estirpes do subtipo A que circulam na África Oriental, sendo sugerido a prevalência e circulação de um recombinante A/G IbNG-like [CRF AG (IbNG)] na Guiné-Bissau (Andersson *et al*, 1999). Mais recentemente, um estudo de Osmanov *et al*. (2002), mostrou uma vez mais, que as infeções por VIH-1 eram maioritariamente classificadas como subtipo A, e que cerca de 31% eram infeções por recombinantes CRF02_AG (Anexo 1). A prevalência da CRF02_AG aumentou progressivamente na Guiné Bissau podendo representar cerca de 50-60% das infeções por VIH-1 no país (Esbjornsson *et al*, 2011; Palm *et al*, 2013). Na população deste país foram também identificadas infeções pelo sub-subtipo A3 numa proporção de 20-30%. (Hemelaar *et al*, 2006; Esbjornsson *et al*, 2011; Palm *et al*, 2013). Em consequência da elevada frequência destes vírus na população foi identificada uma nova variante com origem na recombinação entre a CRF02_AG e o sub-subtipo A3, e que eventualmente poderá estar em crescimento no país (Palm *et al*, 2013; Palm *et al*, 2015). A infeção pelos subtipos B e D parece ser quase inexistente na Guiné Bissau (Esbjornsson *et al*, 2011).

Um estudo realizado no laboratório Nacional de Referência do VIH no INSA, em amostras colhidas a indivíduos atendidos no hospital de Cumura em Bissau revelou uma prevalência elevada da CRF02_AG e a presença de algumas variantes com genomas de mosaico complexos, resultado da recombinação entre diferentes subtipos virais, incluindo C e J (Vaz Sena, 2010). Neste estudo não foram detetadas infeções por subtipos A puros (dados não publicados). Sequências VIH-1 submetidas à base de dados de isolados da Guiné- Bissau revelam a presença dos subtipos A e C, da forma recombinante CRF02_AG e de variantes recombinadas entre a CRF02_AG e os sub-subtipos A1 e A3 no país (Figura 1.4).

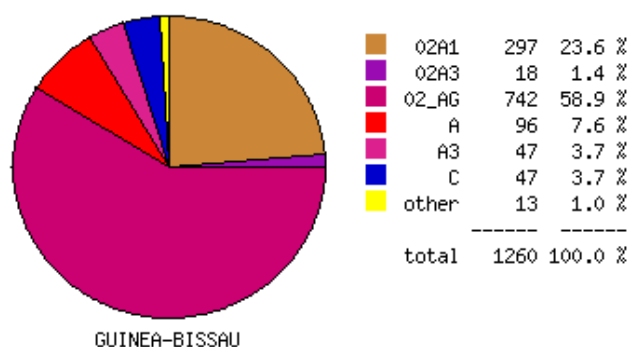


Figura 1.4. Frequência de Subtipos e CRF do VIH-1 existente na base de dados em circulação na população da Guiné-Bissau. (http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp?organism=HIV-1®ion=®ion=gw&form=all&prev_region=pt&path=%2Fhome%2Fhttpd%2Fhtdocs%2Fcomponents%2Fsequence%2FHIV%2Fgeo)

1.4.2 Em Portugal

Portugal é o país com a incidência de VIH-1 mais elevada comparativamente com outros países Europeus, embora na última década se tenha verificado um decréscimo da infeção no país (Palma *et al*, 2007; Carvalho *et al*, 2015; Departamento de Doenças Infecciosas do INSA & Programa Nacional para a Infeção VIH/SIDA, 2015). Em Portugal existe um elevado número de imigrantes, principalmente originários de países Africanos, assim como emigrantes Portugueses no estrangeiro por longos períodos de tempo, mas que acabam geralmente por regressar ao país (Esteves *et al*, 2002; Abecasis *et al*, 2013). Desta forma, verifica-se um elevado potencial para a introdução de diferentes subtipos e formas recombinantes em Portugal, sendo que os estudos mais recentes revelam uma prevalência elevada de subtipos não-B no país, refletindo a epidemiologia da infeção nos diferentes países originários da população migrante (Perrin *et al*, 2003). Assim, em contraste com a maioria dos países Europeus em que predomina a infeção pelo subtipo B, Portugal possui um perfil molecular da infeção VIH-1 único na Europa (Esteves *et al*, 2002; Carvalho *et al*, 2015).

Os primeiros estudos de epidemiologia molecular realizados em Portugal revelaram uma frequência de 50,2% de infeção pelo subtipo B do VIH-1 face a uma proporção equivalente (49,8%) de infeções com subtipos não-B. No grupo de subtipos não-B, observou-se uma prevalência de 21,7% de indivíduos infetados com o subtipo G, 17,5% com o subtipo A, sendo também identificada a infeção por outros subtipos virais (C, D, F, H e J) (Esteves *et al*, 2002).

Estudos mais recentes indicam já uma proporção maioritária (60,8%) de infeções por subtipos não-B em Portugal (Abecasis *et al*, 2013).

Num estudo realizado em 2007 foi descrito o predomínio do subtipo B (41,7%) e do subtipo G (29,4%) em Portugal (Palma *et al*, 2007), no entanto, foi descrito atualmente um decréscimo da infeção pelo subtipo B (26%) (Carvalho *et al*, 2015). No início da epidemia, a elevada frequência dos subtipos B e G na população portuguesa deu origem a um recombinante CRF14_BG (Anexo 1). Este recombinante foi isolado pela primeira vez em Espanha. Porém, a elevada diversidade genética de isolados identificados em Portugal, sugere que a CRF14_BG tenha origem portuguesa (Bártolo *et al*, 2011; Carvalho *et al*, 2015). No entanto, atualmente a forma recombinante com maior prevalência no país é a CRF02_AG, originária e amplamente distribuída no continente Africano (Palma *et al*, 2007; Abecasis *et al*, 2013).

Em contraste, a infeção pelo subtipo C predominante no continente Africano, tem uma frequência reduzida no nosso país (14,5%), assim como as infeções pelo subtipo A1 (4,2%) (Carvalho *et al*, 2015). Na consulta da base de dados de sequências VIH-1 de *Los Alamos*, verificamos que em Portugal foram identificadas infeções pelos subtipos A, B, C, D, G e F e por formas recombinantes AG e BG (http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp?organism=HIV-1®ion=world®ion=pt&form=all&prev_region=world&path=%2Fhome%2Fhttpd%2Fhtdocs%2Fcomponents%2Fsequence%2FHIV%2Fgeo).

1.5 Genes *env*, *nef* e *pol* do VIH-1

1.5.1 Gene *env* e glicoproteína gp120 (SU)

O gene *env* codifica um precursor glicoproteico, a gp160 que é transportada para o complexo de Golgi onde é clivada por uma protease celular originando a gp120 (SU) e a gp41 (TM). Posteriormente, estas glicoproteínas localizadas junto da membrana celular contribuem para a fusão com outras células (formação de sincícios), ou para a formação de novas partículas virais (Kwong *et al*, 2000; Arrildt *et al*, 2012; Matos *et al*, 2014). Por sua vez, incorporadas nos viriões, vão desempenhar papel essencial na ligação ao recetor e coreceptor da nova célula a infectar e conduzir à fusão do invólucro viral com a membrana da célula.

A glicoproteína 120 é composta por dois grandes domínios, interno e externo, contendo cinco regiões conservadas (C1-C5) intercaladas por cinco domínios altamente variáveis (V1- V5) (Figura 1.5). Os domínios variáveis são particularmente expostos, sofrendo alterações em consequência da pressão seletiva, quer pelos anticorpos contra o VIH-1, quer pela ação dos inibidores terapêuticos (ex. antirretrovirais), podendo deste modo ocorrer a evasão do vírus ao sistema imunitário. (Foda *et al*, 2001; Clapham & McKnight, 2002).

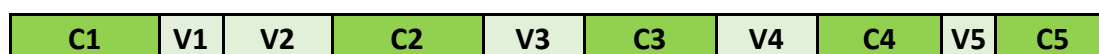


Figura 1.5. Representação esquemática dos domínios internos (C) e externos (V) da gp120 de VIH-1.
Adaptado de Khayat *et al*, 2013

Numa fase inicial, a gp120 liga-se à molécula CD4 e há uma alteração na conformação da glicoproteína do vírus promovendo a sua interação com os recetores das quimiocinas, CCR5 e CXCR4 (Bagaya *et al*, 2015). A escolha do recetor primário a ser utilizado para a entrada do vírus na célula deve-se à capacidade de formar sincícios, assim como ao respetivo tropismo M (R5) ou T (X4) das variantes virais (Cho *et al*, 1998). As variantes X4 infetam os macrófagos e possuem a capacidade de formar sincícios (SI) e de usar o coreceptor CXCR4, enquanto que as variantes R5, associadas à infeção das células T, não possuem capacidade de formar sincícios (NSI) e apresentam elevada afinidade com o coreceptor CCR5 (Resch *et al*, 2001; Jesen *et al*, 2003; Arrildt *et al*, 2012). A utilização dos co-recetores pode estar relacionada com o tempo de infeção do indivíduo, pelo que a escolha do co-recetor CCR5 é uma característica comum das variantes virais da fase inicial da infeção, antes desta se estabelecer. Por outro lado, a utilização do co-recetor CXCR4 observa-se em fases mais avançadas da infeção (Azevedo-Pereira & Costa-Santos, 2008)

Diferentes resíduos ou motivos da proteína Env são essenciais na progressão da infeção. As posições 11 e 25 do *loop* V3 são muito conservadas nesta região. Na primeira posição todos os isolados NSI apresentam maioritariamente aminoácidos com cargas neutras ou negativas, respetivamente S/G ou E/D, enquanto na posição 25 os isolados SI apresentam geralmente cargas positivas (Xiao *et al*, 1998)

Foi identificado o motivo S/GxxxGPGxxxxxxxxE/D, conservado entre os isolados de alguns subtipos A, B, C, D, E e F do VIH-1, que utilizam o CCR5 como principal recetor para a entrada do vírus na célula (Xiao *et al*, 1998; Resch *et al*, 2001; Cormier & Dragic, 2002). Também as duas cisteínas (posições 1 e 35) são altamente conservadas, formando uma ponte de dissulfureto entre si, essencial para preservar a conformação do *loop* V3 de Env (Resch *et al*, 2001).

1.5.2 Gene *nef* e Proteína Nef

O gene *nef* codifica a proteína Nef que possui múltiplas funções, destacando-se a regulação negativa de moléculas CD4 e de moléculas MHC-I e II (complexo principal de histocompatibilidade de classe I e II) na superfície da célula, assim como exerce interferência com proteínas celulares envolvidas nas vias de transdução de sinal e no aumento da infeciosidade viral (Aiken & Trono, 1995; Greenway *et al*, 2003; Li *et al*, 2005).

Após a tradução da proteína Nef, esta sofre meristilação no terminal amina com posterior clivagem pela protease viral, processos essenciais para a sua ligação à membrana celular. A meristilação ocorre no resíduo de glicina do motivo MGxxxS, altamente conservado em todas as variantes de VIH-1 (Geyer *et al*, 2001; Kirchhoff *et al*, 2008). A clivagem da proteína ocorre entre o triptofano 57 (W₅₇) e a leucina 58 (L₅₈) (Chen *et al*, 1998), separando a proteína em dois domínios distintos, respetivamente a âncora e o core (Geyer *et al*, 2001).

Na regulação negativa de moléculas CD4, a proteína Nef também pode sofrer fosforilação, estando neste processo associado o motivo RPMTYK77 (resíduo de treonina). Na interferência de Nef nos mecanismos de regulação celular estão principalmente envolvidos os resíduos WL57, E59, A60, G95, L97, R106, I109 e L110 e o motivo FDP₁₂₁, respetivamente associados à ligação de Nef a moléculas CD4 e a proteínas essenciais ao tráfego celular (Kirchhoff *et al*, 2008; Geyer *et al*, 2001).

A conservação de outros resíduos, EE₁₅₄ e M₂₀, e de outros motivos tais como tetraprolina (PxxPxxPxxP), a dileucina (LL₁₆₅) e o agrupamento ácido (EEEE₆₂), são considerados essenciais no contexto de uma combinação funcional, estrutural e bioquímica na regulação negativa da célula infetada exercida pela proteína Nef (Piguet *et al*, 1999; Basmaciogullari & Pizzato, 2014) (Figura 1.6).

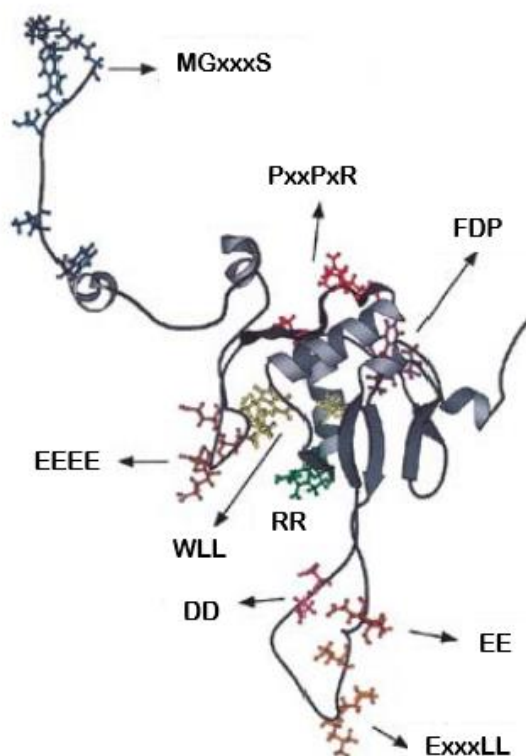


Figura 1.6. Principais motivos estruturais e funcionais da proteína Nef do VIH-1. Adaptado de Geyer et al, 2001

1.5.3 Gene *pol* e as Proteínas PR e RT

1.5.3.1 Protease

A análise cristalográfica aplicada ao estudo da protease revelou que a enzima apresenta na sua zona central um único centro ativo com duas estruturas móveis (*flaps*) (Verveka *et al*, 2003; Hornak & Simmerling, 2007). Estas estruturas abrem e fecham sobre o centro ativo da enzima (Tomasselli & Heinrikson, 2000). No desempenho das suas funções, a protease possui o motivo D₂₅T₂₆G₂₇, da tríade catalítica altamente conservado (Hou *et al*, 2008). O motivo estrutural I₄₇GGIGG₅₃ encontra-se presente num dos *flaps* assumindo um papel fundamental na ligação da enzima ao substrato (Tomasselli & Heinrikson, 2000).

O centro ativo da protease é constituído por vários resíduos altamente conservados (R₈, D₂₅, G₂₇, A₂₈, D₂₉, G₄₉, V₈₂ e I₈₄) importantes para a manutenção da atividade da enzima e constituem um ótimo alvo para os agentes inibidores da protease (Tomasselli & Heinrikson, 2000; Hou *et al*, 2008) (Figura 1.7).

A protease é essencial para o ciclo replicativo do VIH, uma vez que é indispensável para a maturação e infecciosidade dos viriões sendo por isso um importante alvo terapêutico (Hou *et al*, 2008).

Em condições de pressão terapêutica a protease pode sofrer alterações que permitem manter as suas funções, e simultaneamente, reduzir a suscetibilidade aos inibidores da protease administrados no tratamento da infeção. (Tomasselli & Heinrikson, 2000; Todd *et al*, 2000).

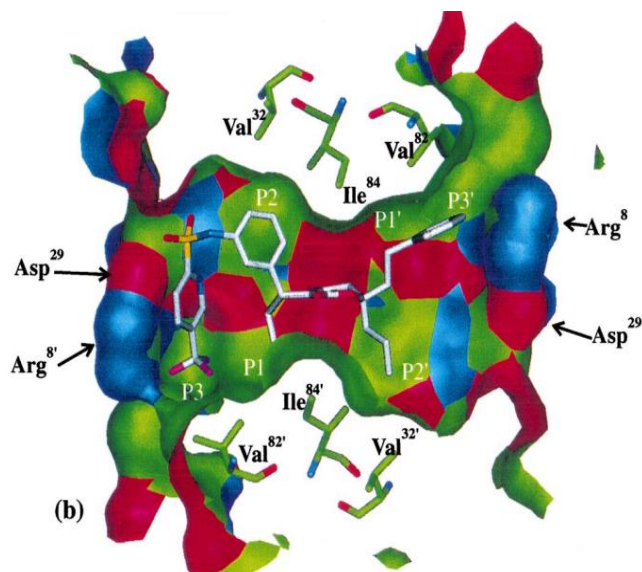


Figura 1.7. RX da análise cristalográfica do centro ativo da região da protease do VIH-1, com os principais resíduos altamente conservados desta enzima. Adaptado de Tomasselli & Heinrikson, 2000.

1.5.3.2 Transcriptase Reversa

A transcriptase reversa possui duas importantes funções enzimática: A polimerização de ADN, consistindo na síntese de cadeias de ADN a partir do modelo de ADN ou ARN, e a atividade da ribonuclease H (RNase H) que degrada o ARN viral presente na cadeia híbrida de ARN/ADN (Mulky *et al*, 2005; Sarafianos *et al*, 2009). Estas duas atividades são essenciais para a conversão do genoma ARN de cadeia simples em ADN de cadeia dupla (Gutiérrez-Rivas & Menéndez-Arias, 2001; Sluis-Cremer *et al*, 2003; Sarafianos *et al*, 2009).

A RT é constituída por duas subunidades, a p51 (menor) e a p66 (maior). A primeira tem um papel estrutural e deriva da p66 por remoção proteolítica do domínio RNase H, e a segunda, apresenta os domínios catalíticos, polimerase e a RNase H (Tachedjian *et al*, 2003; Mulky *et*

al, 2007). O domínio da polimerase assemelha-se à figura de uma mão humana e divide-se em subdomínios (*fingers*, *palm* e *thumb*) (Gutiérrez-Rivas & Menéndez-Arias, 2001).

O local de ligação do ARN viral permite o contato com o centro ativo da polimerase e RNase H, sendo o primeiro constituído por três carboxilatos catalíticos contendo os resíduos D₁₁₀, D₁₈₅ e D₁₈₆. Os resíduos de aspartato na posição 185 e 186 constituem motivo altamente conservado YMDD (Sharma *et al*, 2005). Adicionalmente, os resíduos W229, G231 e L234 que mantêm a conformação do centro ativo da polimerase e os resíduos K₆₅, R₇₂, Y₁₁₅ e Q₁₅₁ que correspondem ao local de ligação aos dNTP são observados conservados (Gutiérrez-Rivas & Menéndez-Arias, 2001; Mulky *et al*, 2007). Alterações de aminoácidos nessas posições levam à perda de atividade ou inativação total da enzima (Gutiérrez-Rivas & Menéndez-Arias, 2001; Sarafianos *et al*, 2009).

1.6 Inibidores da Transcriptase Reversa e Protease

Ao longo do tempo os diferentes regimes de antirretrovirais administrados no tratamento influenciaram as taxas de mutação e seleção de variantes com mutações de resistência (Chen *et al*, 2005). O tratamento com antirretrovirais de baixa eficácia ou associado a uma baixa adesão dos doentes conduziu à seleção variantes resistentes, que por sua vez, podem ser transmitidas a novos indivíduos (Shafer *et al*, 2007).

Os primeiros fármacos desenvolvidos contra a infeção VIH-1 foram inibidores que tinham como alvo a transcriptase reversa. Estes inibidores incluem a classe dos inibidores de nucleósidos da transcriptase reversa (NRTI) e dos inibidores não nucleósidos da transcriptase reversa (NNRTI). Apenas, no final do século passado surgiram os inibidores da protease como o segundo alvo terapêutico no combate à infeção VIH-1. Os inibidores da protease (PI) tem como alvo inibir a atividade enzimática da protease do VIH-1, que é suficiente para evitar os eventos de clivagem em *gag* e *gag-pol* resultando na produção de partículas não infecciosas (Fun *et al*, 2011; Adamson, 2012). Estas classes de inibidores enzimáticos, da RT e da PR são as classes mais usadas no tratamento da infeção VIH (Chen *et al*, 2005; Adamson, 2012).

A terapia antirretroviral de elevada eficácia (HAART) combina a utilização de inibidores da protease com inibidores da transcriptase reversa (nucleósidos e não nucleósidos), sendo

atualmente o regime terapêutico frequentemente administrado no tratamento e na prevenção da transmissão mãe-filho do VIH-1 (Adamson, 2012; Lv *et al*, 2015). A sua eficácia no bloqueio do ciclo de replicação do vírus associada à elevada adesão do doente ao tratamento tem conduzido ao decréscimo da mortalidade e morbilidade e ao aumento da esperança de vida dos indivíduos infetados (Pokorná *et al*, 2009). Porém, o aparecimento de mutações de resistência é reconhecido como a causa de insucesso no tratamento da infeção VIH/SIDA.

- Os NRTI são pró-fármacos que requerem uma fosforilação por cinases para se tornarem ativos. Estes inibidores são similares a desoxirribonucleósidos que ocorrem de forma natural e a sua introdução na cadeia de ADN leva à supressão da síntese de ácido nucleico (Sluis-Cremer *et al*, 2004; Arts & Hazuda, 2012). Existem descritos na literatura várias mutações que podem reverter a terminação da cadeia de ADN nomeadamente, M41L, D67N, K70R, L210W, T215Y/F ou K219Q/E. Adicionalmente, encontram-se também descritas as mutações K65R, L74V, Q151M, M184V/I associadas à resistência aos NRTI (Arts & Hazuda, 2012).
- Os NNRTI inibem a atividade da RT do VIH-1 pela sua ligação e consequentemente uma alteração na conformação espacial no centro ativo da polimerase (Sluis-Cremer & Tachedjian, 2008). As substituições de aminoácidos associadas à resistência ocorrem ao nível da região de ligação aos NNRTI (Sluis-Cremer *et al*, 2004; Arts & Hazuda, 2012). As mutações de resistência mais comuns são a K103N e Y181C (Arts & Hazuda, 2012).
- A resistência aos PI ocorre de forma gradual e resulta de combinações de mutações múltiplas (Adamson, 2012). As primeiras mutações a surgir são denominadas primárias e conduzem à aquisição de resistência aos inibidores da protease. As mutações aos PI descritas incluem as seguintes alterações: D30N, G48V, I50L/V, V82A/F/G/C/T, I84V ou L90M (Adamson, 2012; Arts & Hazuda, 2012; Flynn *et al*, 2015).

2. Enquadramento e Objetivos do Estudo

No Laboratório Nacional de Referência VIH foi realizado um estudo de prevalência do VIH a partir da análise amostras de indivíduos naturais e residentes na Guiné-Bissau, colhidas no Hospital de Cumura durante o ano de 2008. As amostras VIH-1 positivas foram submetidas a amplificação e sequenciação da região V3-V5 e do gene *nef* para a classificação molecular dos vírus em circulação no grupo estudado. Neste estudo não foram identificados subtipos puros de VIH-1 e a maioria dos vírus foram classificados em recombinantes CRF02_AG. Foram também identificadas variantes com genoma mosaico (potencialmente URF) resultado de possíveis eventos de recombinação entre a CRF02_AG e outros subtipos de VIH-1 (subtipos C e J). A maioria das sequências *env* e *nef* do VIH-1 obtidas neste estudo encontram-se disponíveis na base de dados de sequências de VIH de *Los Alamos* e foram utilizadas como sequências de referência no desenvolvimento do presente trabalho, que agora pretende avaliar a diversidade do VIH-1 em indivíduos Guineenses residentes em Portugal.

Assim, foram estabelecidos dois objetivos gerais:

- Avaliar a diversidade do VIH-1 no grupo de Guineenses residentes em Portugal
- Comparar a diversidade genética encontrada com a do grupo de Guineenses residentes na Guiné-Bissau

Deste modo, foram estabelecidos os seguintes objetivos específicos:

- a) Amplificar e sequenciar as regiões *env*, *nef* e *pol* (PR e RT) do VIH-1, a partir de amostras de ADN excedentes no laboratório, colhidas a indivíduos de naturalidade Guineense residentes em Portugal;
- b) Realizar a análise filogenética para classificação do VIH-1 e conhecer os subtipos e os recombinantes do VIH-1 na amostra estudada;
- c) Analisar os principais motivos estruturais e funcionais de Env e Nef;
- d) Realizar a previsão do uso de coreceptores CCR5 e CXCR4;
- e) Avaliar as distâncias nucleotídicas das sequências de *env* e *nef*;
- f) Analisar as mutações de resistência aos inibidores da TR e da PR.

Com o desenvolvimento deste trabalho, pretendeu-se adquirir um melhor conhecimento sobre a epidemiologia e evolução da infeção VIH-1 no grupo de Guineenses residentes em Portugal.

3. Material e Métodos

3.1 População estudada

A população estudada corresponde a uma amostra de conveniência constituída por indivíduos do sexo feminino infetados com VIH-1 de naturalidade Guineense e residentes em Portugal. As amostras sanguíneas foram recebidas entre 2012 e 2015, no Laboratório Nacional de referência VIH e Hepatites B e C, no âmbito do protocolo laboratorial implementado para estudo da transmissão mãe-filho do VIH. Amostras de ADN proviral do VIH-1 excedente, foram conservadas a -80°C até à sua utilização. O presente estudo foi retrospectivo, anónimo e não relacionável. A recolha dos dados demográficos (sexo, idade, naturalidade) e a informação sobre o cumprimento de regimes de terapia antirretroviral de prevenção (Sim ou Não) foi realizada por consulta dos registos numa base de dados anonimizada disponível no laboratório.

3.2 Extração de ADN proviral do VIH-1

A extração de ADN proviral do VIH-1 foi realizada a partir de células mononucleares de sangue periférico (CMSP), recorrendo ao kit QIAmp DNA Blood (QIAGEN, Studio City, USA) seguindo as normas de utilização do fabricante.

O procedimento de extração de ADN proviral VIH-1 foi realizado apenas para uma pequena parte das amostras analisadas. Na maioria dos casos as amostras de ADN encontravam-se armazenadas e conservadas no laboratório.

3.3 Caracterização dos *primers* utilizados no estudo

Para a amplificação das regiões alvo (*env*, *nef* e *pol*) foram utilizados diferentes pares de *primers* para a deteção do VIH-1. Com exceção da região *pol* em que foram utilizados quatro *primers* externos na primeira reação de amplificação, nas regiões *env* e *nef* foram usados dois *primers* externos. Em todas as segundas reações de amplificação (*env*, *nef*, PR e RT) foi usado um par de *primers* internos.

As regiões de *env* e *nef* do VIH-1 foram amplificadas por *nested*-PCR utilizando respetivamente os *primers* descritos por Delwart *et al.* (1997) e Parreira *et al.* (2005). Para a região *pol* foram utilizados os *primers* descritos por Steegen *et al.* (2006) e Videira e Castro (2010). As diferentes sequências dos *primers* selecionados para amplificação das regiões *env* e *pol* e o gene *nef* do

VIH-1 encontram-se descritas na Tabela 3.1, sendo a respetiva numeração da posição dos nucleótidos efetuada com base na sequência de referência HXB2 do VIH-1 retirada da base de dados *Los Alamos*, (*GenBank*, número de acesso:K03455).

Todos os *primers* utilizados neste estudo foram adquiridos na forma liofilizada e a sua hidratação foi efetuada em “sala limpa” pela adição de água ultrapura para obter os respetivos *stocks* a uma concentração 100 µM. Posteriormente, cada um deles foi diluído em soluções de trabalho de 10 µM, para serem usadas nas reações de amplificação e diluídos a 1µM para a reação de sequenciação.

Tabela 3.1. Primers utilizados nas misturas de reação para amplificação das diferentes regiões genómicas do VIH-1

Região alvo a amplificar	Primer	Sequência 5'-3'	Posição em HXB2	Fragmento esperado (em pb)
<i>env</i> (1)	ED5	ATGGGATCAAAGCCTAAAGCCATGTG	6557 - 6582	685 pb
	ED12	AGTGCTTCCTGCTGCTGCTCCCAAGAACCAAG	7782- 7811	
	ES7	TGTAAAACGACGGCCAGTCTGTAAATGGCAGTCTAGC	6983 - 7021	
	ES8	CACCTTCTCCAATTGTCCCTCA	7648 - 7668	
<i>nef</i> (2)	H1NefOf	GTGCCTCTTCAGCTACCACGC	8513 - 8533	771 pb
	H1NefOr	AGCATCTGAGGGTTAGCCACT	9488 - 9508	
	H1Neflf	TGGACAGATAGGGTTATAGAA	8697 - 8717	
	H1Neflr	CACCTCCCCTGGAAAGTCCCC	9448 - 9468	
<i>pol</i> (3)	Gag2	CAATGAGGAAGCTGCAGAATGGG	1404-1426	
	PR1	GAAATGATGACAGCATGYCAGG G	1822-1844	
	RT137	TTCTGTATRTCATTGACAGTCCAGC	3301-3325	
	RT3303	TAAYTTYTGATATRTCATTGAC	3309-3329	
Protease (4)	PR3	CCAGAGCCAACAGCCCCACCA	2146-2166	459 pb
	PR4	GGGCCATCCATTCTGGCTTTA	2584-2605	
Trasncriptase reversa (5)	RT1	CCAAAAGTTAAACAATGGCCATTGACAGA	2604-2632	649 pb
	RT4	TGGAGTTCATAWCCCATCCAAAG	3231-3253	

(1) Delwart et al. (1993)

(2) Parreira et al. (2005)

(3), (4) e (5) Steegen et al. (2006) e Videira e Castro (2010)

3.3.1 Primers para a amplificação das regiões env, nef, PR e RT do VIH-1

a) *Primers env e nef*

A primeira reação de amplificação da região *env* do VIH-1 foi efetuada com os *primers* externos ED5 e ED12 (direto e reverso). Na segunda reação de amplificação foram utilizados os *primers* internos ES7 e ES8 (direto e reverso). Para a primeira amplificação do gene *nef* do VIH-1 foram utilizados os *primers* externos H1NefOf e H1NefOr, tendo sido utilizado na segunda reação de amplificação os *primers* internos H1NefIf e H1NefIr.

b) *Primers PR e RT (pol)*

Na primeira reação de amplificação da região *pol* do VIH-1 foram adicionados na mistura de reação quatro *primers* externos, dois diretos (Gag2 e PR1) e dois reversos (RT137 e RT3303). A sua utilização em simultâneo na mistura conduziu à amplificação de vários fragmentos de ADN que serviram de matriz para uma segunda reação de amplificação. Nesta segunda reação foram utilizados de forma individualizada os *primers* internos RT1 e RT4 (direto e reverso) para a amplificação da região RT e os *primers* internos PR3 e PR4 (direto e reverso) para amplificação da região PR.

3.4 Condições das reações de amplificação do VIH-1

A primeira reação de amplificação, de todas as regiões genómicas alvo foi realizada com o sistema *illustra™ PuRe Taq Ready-to-go PCR Beads* (GE Healthcare, England). Os microtubos possuem reagentes liofilizados em forma de uma pequena esfera. Estas esferas são compostas por vários componentes, entre os quais, o tampão e estabilizadores da reação, BSA (*bovine serum albumine*), dNTPs, MgCl₂ e cerca de 2,5U de enzima *PuReTaq* DNA polimerase. A reconstituição das esferas foi feita para um volume total de 25 µl com uma concentração de 0,4 mM de cada par de *primers* externos de *env* e de *nef*, e uma concentração de 0,2 mM de cada um dos quatro *primers pol* externos. A cada microtubo foram adicionados 6 µl de amostra de ADN e posteriormente submetidos aos programas experimentais descritos na tabela 3.2.

Na segunda reação de amplificação foi utilizada a enzima *AmpliTaq Gold* (Applied Biosystems, Foster City, USA). No entanto, as concentrações dos diferentes componentes das misturas e as condições de amplificação foram distintas para cada uma das regiões genómicas amplificadas.

a) Segunda reação de amplificação da região *env*

A segunda reação de amplificação da região *env* foi efetuada num volume total de 100 µl contendo a seguinte concentração de reagentes: 1x de PCR Buffer, 1,5 mM de MgCl₂, 0,2 mM de dNTPs, 0,4 mM de cada um dos *primers* e 1,25 U da enzima AmpliTaq Gold. Posteriormente Foi adicionado 1 µl de produto amplificado na primeira reação e os microtubos foram colocados num termociclador com o programa de amplificação descrito na Tabela 3.2.

Tabela 3.2. Programas de amplificação (1ª e 2ª reação de PCR) utilizados para a amplificação da região *env* do VIH-1

1ª reação de PCR	Desnaturação inicial	95°C, 5 min
	Amplificação	95°C, 30seg; 45°C, 45 seg; 72°C, 1 min; x10
		95°C, 30seg; 50°C, 45 seg; 72°C, 1 min; x30
	Extensão final	72°C, 7 min
2ª reação de PCR	Desnaturação inicial	95°C, 5 min
	Amplificação	95°C, 30seg; 45°C, 45 seg; 72°C, 1 min; x5
		95°C, 30seg; 50°C, 45 seg; 72°C, 1 min; x15
		95°C, 30seg; 55°C, 45 seg; 72°C, 1 min; x20
	Extensão final	72°C, 7 min

b) Segunda reação de amplificação do gene *nef* do VIH-1

A mistura de reação para a segunda amplificação do gene *nef* foi efetuada a partir de um volume total de 50 µl com a concentração final dos seguintes reagentes: 1x de PCR Buffer, 1,5 mM de MgCl₂, 0,2 mM de dNTPs, 0,4 mM de cada *primer* e 1,25 U da enzima AmpliTaq Gold (*Applied Biosystems, Foster City, USA*). Foram adicionados à mistura de reação 2 µl de produto amplificado na primeira reação de amplificação e os microtubos colocados no termociclador com o programa descrito na tabela 3.3.

Tabela 3.3. Programa de amplificação (1ª e 2ª reação de PCR) do gene de *nef* do HIV-1

1ª reação de PCR	Desnaturação inicial	95°C, 5 min
	Amplificação	95°C, 45 seg; 38°C, 45 seg; 72°C, 45 seg; x3
		95°C, 30seg; 50°C, 45 seg; 72°C, 45 seg; x40
	Extensão final	72°C, 7 min
2ª reação de PCR	Desnaturação inicial	95°C, 5 min
	Amplificação	95°C, 30seg; 50°C, 45 seg; 72°C, 45 seg; x15
		95°C, 30seg; 55°C, 1 min; 72°C, 1 min; x30
	Extensão final	72°C, 7 min

c) Segunda amplificação da região *pol* (PR e RT)

A segunda reação de PCR correspondente à amplificação de PR e RT, efetuou-se num volume final de 50 µl com misturas contendo os *primers* PR e RT em separado, a que se adicionaram reagentes nas seguintes concentrações: 1x de PCR Buffer, 1,5 mM de MgCl₂, 0,2 mM de dNTPs e 1,25 U da enzima AmpliTaq Gold. A estas misturas de reação foi adicionado 1 µl do produto amplificado na primeira reação de amplificação. O programa de amplificação utilizado encontra-se descrito na Tabela 3.4.

Tabela 3.4. Programa de amplificação (1ª e 2ª reação de PCR) utilizado na amplificação de PR e de RT do VIH-1

1ª reação de PCR	Desnaturação inicial	95°C, 5 min
	Amplificação	95°C, 45 seg; 45°C, 1 min; 72°C, 1 min; x10
		95°C, 45 seg; 50°C, 1 min; 72°C, 1 min; x25
		95°C, 45 seg; 55°C, 1 min; 72°C, 1 min; x20
	Extensão final	72°C, 7 min
2ª reação de PCR	Desnaturação inicial	95°C, 5 min
	Amplificação	95°C, 45 seg; 50°C, 1 min; 72°C, 1 min; x20
		95°C, 45 seg; 55°C, 1 min; 72°C, 1 min; x20
	Extensão final	72°C, 7 min

3.5 Visualização de produtos amplificados (2ª amplificação)

Os produtos de PCR obtidos na segunda reação de amplificação foram visualizados através de uma eletroforese em gel de agarose a 1,4% em tampão TAE 1x, sendo o gel contrastado com corante SyberSafe (Invitrogen, California, USA). A cada 5 µl do produto amplificado foi

adicionado 3 µl de azul bromofenol e a mistura colocada em cada poço do gel. Na corrida eletroforética foi aplicado o marcador de massa molecular 1 Kb Plus DNA Ladder (Invitrogen, California, USA) de forma a verificar se a massa molecular do fragmento obtido nas amostras era semelhante ao fragmento esperado. A corrida eletroforética decorreu durante 1 hora e 30 minutos a cerca de 120 V (volts) e os fragmentos de ADN do gel foram observados num transiluminador e fotografados sob luz UV usando o sistema Gel-Doc (Bio-Rad, USA).

3.6 Purificação e Sequenciação dos produtos de PCR

Todas as amostras contendo o fragmento amplificado de massa molecular esperada (685 pb para *env*, 771 pb para *nef*, 459 para PR e 649 para RT) foram purificados de forma a eliminar os restos de *primers* ou outros constituintes não consumidos nas reações de amplificação e desta forma evitar a sua interferência na reação de sequenciação. A maioria das amostras foi purificada com recurso ao kit *ExoSap- IT USB* (QIAGEN, Valencia, CA), no entanto, em algumas amostras a purificação foi efetuada com o kit *QIAquick® PCR Purification* (QIAGEN, Valencia, CA).

3.6.1 Purificação com o kit ExoSap- IT USB

No procedimento da purificação dos produtos amplificados foi utilizado o kit *ExoSap- IT USB*. Este método enzimático consiste na adição de 2 µl da enzima *ExoSap* a cada 5 µl de produto de PCR. Esta mistura foi colocada num termociclador e submetida a uma incubação a 37°C seguida de uma segunda incubação a 80°C para a inativação da enzima, ambas com duração de 15 minutos.

3.6.2 Purificação com o kit QIAquick® PCR Purification

Nas amostras em que foi necessário realizar a purificação dos produtos amplificados com o kit *QIAquick® PCR Purification* (QIAGEN, Valencia, CA), adicionou-se 125 µl de tampão PB a 25 µl de produto amplificado (1:5). Esta mistura foi colocada numa coluna, adaptada a um tubo coletor de 2 ml e centrifugada a 10000 rpm durante 1 minuto de forma a ocorrer uma ligação entre o ADN e a coluna. Após a centrifugação, o filtrado foi descartado e procedeu-se a uma lavagem do ADN fixado na coluna, adicionando 750 µl de tampão PE e nova centrifugação durante 1 minuto a 10000 rpm. O filtrado foi descartado e feita nova centrifugação em iguais condições. Para eluir o ADN presente na coluna foi adicionado cerca de 30 µl de tampão EB

(10 mM Tris-Cl, pH 8,5) com incubação à temperatura ambiente, de um minuto seguindo-se uma centrifugação para novo tubo coletor. O eluído purificado foi transferido para o tubo com tampa de rosca, para ser conservado até à sua utilização na reação de sequenciação.

3.6.3 Sequenciação dos purificados com o *kit* Big Dye™ Terminator Cycle Sequencing Ready Reaction

Todos os produtos purificados foram submetidos a uma sequenciação nucleotídica, baseada no método de terminação de cadeia (Sanger *et al*, 1977) utilizando o *Kit Big Dye™ Terminator Cycle Sequencing Ready Reaction* (PE Applied Biosystems, Foster City, USA) de acordo com as instruções do fabricante. A mistura de reação de sequenciação foi efetuada em “sala limpa” contendo 1 µl de Big Dye e 0,5 µl de *primer* a 1 µM. Posteriormente numa sala específica para a manipulação de produtos amplificados, adicionou-se entre 1 a 3 µl de produto purificado, dependendo da sua concentração (previamente verificada em gel de agarose), ajustando-se, por fim, o volume de reação com água de forma a perfazer 10 µl. Os tubos foram colocados no termociclador e submetidos ao programa de sequenciação descrito na Tabela 3.5.

Tabela 3.5- Programa de sequenciação nucleotídica.

Programa	
Desnaturação inicial	96°C, 30 seg; x1
Sequenciação	96°C, 30 seg; 50°C, 15 seg; 60°C, 4 min; x25

Após a reação de sequenciação, os tubos foram entregues na Unidade de Tecnologia e Inovação (UTI) do INSA, para serem submetidos a uma nova purificação e posterior corrida das amostras no sequenciador *ABI PRISM 3130 Genetic Analyser* (Applied Biosystems- *life technologies*, Carlsbad, CA). As sequências obtidas foram enviadas ao laboratório para respetiva análise dos cromatogramas e construção de sequências consenso por amostra e por cada região genómica amplificada (*env*, *nef*, PR e RT do VIH-1).

3.7 Análise de Sequências Nucleotídicas

3.7.1 Cromatogramas e Construção de Sequências Consenso

Os cromatogramas foram analisados e as sequências editadas com o auxílio do programa *Chromas Lite 2.4.4*. A construção de sequências consenso para cada amostra e para cada região foi realizada através do recurso ao programa de alinhamento de sequências *BioEdit 7.2.5*.

3.7.2 Inferência Filogenética

Todas as sequências de consenso obtidas para cada amostra e para cada região genómica em estudo foram submetidas à análise prévia de homologia genética através do programa BLAST (*Basic Local Alignment Editor Search Tool*). Esta abordagem inicial ao estudo das sequências permitiu obter uma primeira classificação das sequências VIH-1, e orientou a seleção de sequências de referência a utilizar posteriormente na construção das árvores filogenéticas. Sequências de referência referentes a todos os subtipos representativos da infeção VIH-1 em Portugal e na Guiné-Bissau foram retiradas da base de dados de VIH de *Los Alamos*.

As sequências de consenso de cada amostra e as sequências de referência foram alinhadas entre si, separadamente para cada região em estudo. Foram efetuados pequenos ajustes manuais e realizada a remoção dos espaços no alinhamento, para a posterior construção das árvores filogenéticas. Na construção das árvores filogenéticas utilizou-se o programa *MEGA 7 (Molecular Evolutionary Genetics Analysis)*, selecionando o método de Junção de vizinhos e o modelo de “2 parâmetros de Kimura” (Kimura, 1980), no cálculo da matriz de distâncias nucleotídicas. A avaliação da robustez da topologia da árvore foi feita através do teste de bootstrapping, tendo como base 1000 réplicas aleatórias, sendo considerados estatisticamente válidos os agrupamentos com valores iguais ou superiores a 70%. Na construção das árvores foi usada como *outgroup* a referência SIVcpzMT145.

3.7.3. Distâncias Nucleotídicas

O cálculo das distâncias nucleotídicas da região V3-V5 e do gene *nef* foi baseado nas matrizes de distância obtidas entre os diferentes alinhamentos de sequências selecionadas usando o programa *MEGA 7*, seguindo também, o modelo de “2 parâmetros de Kimura” (Kimura, 1980).

3.8 Sequências aminoacídicas de V3-V5, proteína Nef, PR e RT

As sequências nucleotídicas de consenso de cada amostra e para cada uma das regiões genômicas analisadas foram traduzidas em aminoácidos, com base no código genético universal aplicando o programa *Gene Runner 6.0.28 Beta* e efetuada a análise do grau de conservação dos domínios estruturais e funcionais das diferentes proteínas. Para as sequências da região *env* foi estimada a utilização de co-recetores (CCR5 e CXCR4) a partir do programa WebPSSM (<https://indra.mullins.microbiol.washington.edu/webpssm/>) e determinado o número de potenciais locais de N-glicosilação com a ferramenta bioinformática *N-Glycosite* disponível na base de dados de VIH, *Los Alamos* (<http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html>).

Para as sequências da região pol foram analisadas as mutações associadas a resistência aos inibidores da PR e RT recorrendo ao programa Geno2Pheno (<http://www.geno2pheno.org/>).

4. Resultados

4.1 Caracterização das Amostras Estudadas

A amostra de conveniência estudada incluiu 42 amostras de indivíduos infectados por VIH-1 por via heterossexual, todos do sexo feminino, naturalidade Guineense e residentes em Portugal. A idade destas mulheres variou entre 18 e 40 anos, sendo a média de idades da população de 30,9 anos e a mediana de 31 anos. As amostras sanguíneas foram colhidas entre 2 a 30 dias após o parto e enviadas ao laboratório no contexto do estudo da transmissão mãe-filho do VIH. A maioria destas mulheres (90,5%; n=38) cumpriu regime de terapia antirretroviral durante a gravidez, no entanto, verificou-se que em alguns casos (9,5%; n=4) não foi seguida qualquer medida de prevenção (GP232, GP245, GP249 e GP256). Observou-se que uma mulher que não fez prevenção transmitiu o VIH-1 ao filho (GP256).

4.2 Amplificação das regiões *env*, *nef*, PR e RT de VIH-1

As amostras de ADN foram sujeitas a um *nested*-PCR para a amplificação das regiões *env*, *nef* e *pol* (PR e RT) do VIH-1. A amplificação foi observada no total das 42 (100%) amostras analisadas para a região *env*, em 33 (79%) amostras para *nef*, em 35 (83%) amostras para a região PR e em 37 (86%) amostras para a região RT. Como exemplo ilustrativo foram apresentados na Figura 4.1 os resultados de quatro amostras submetidas a amplificação para as quatro regiões em análise.

Todas as amostras amplificadas foram purificadas e sequenciadas. No entanto, nem todos os cromatogramas apresentaram uma leitura com qualidade suficiente para a construção de sequências consenso. Deste modo, foi possível construir um total de 30 (30/42; 70%) sequências consenso para a região *env*, 27 (27/33; 82%) para *nef*, 24 (24/35; 69%) para PR e 34 (34/37; 92%) para RT, cada uma correspondendo a uma amostra/ indivíduo em estudo.

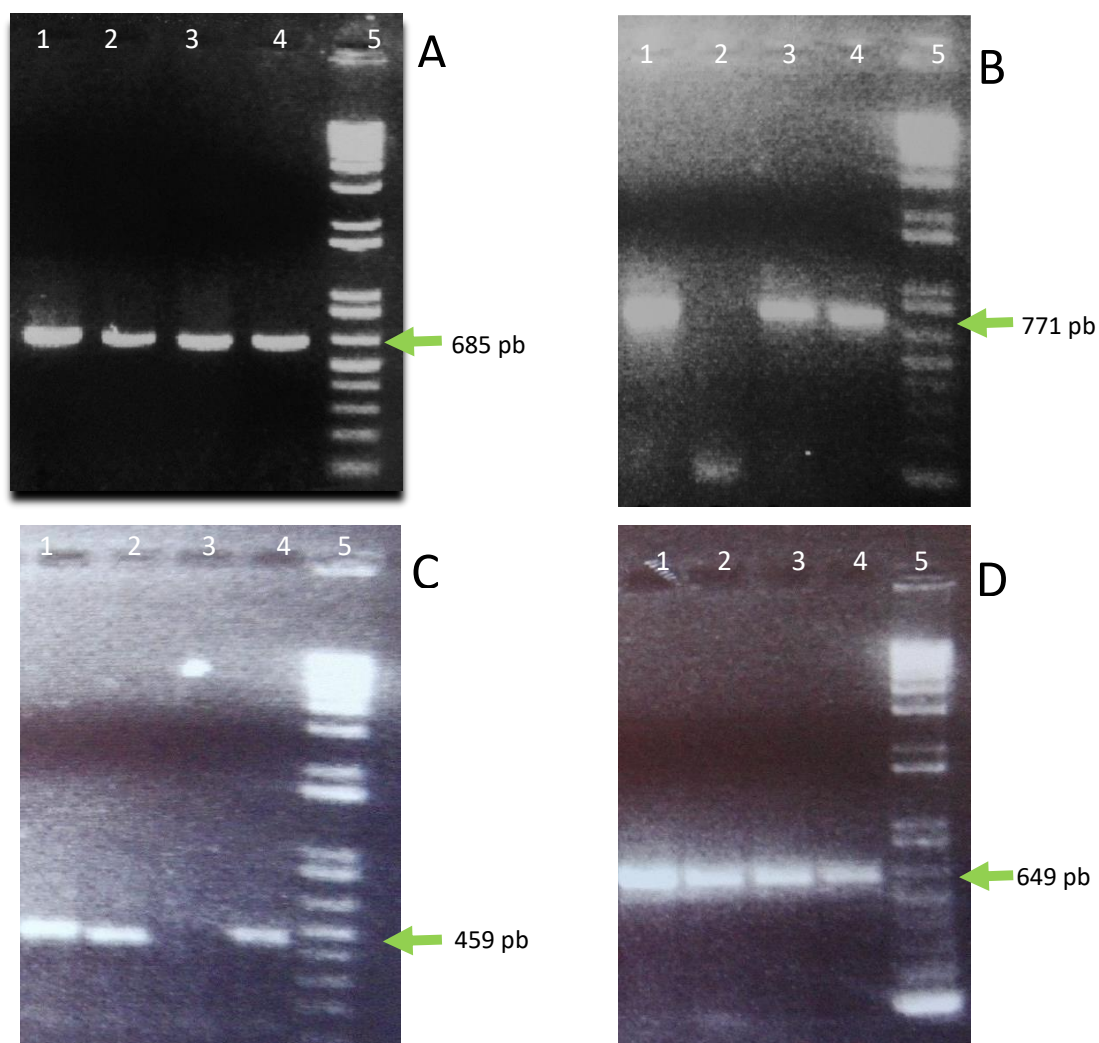
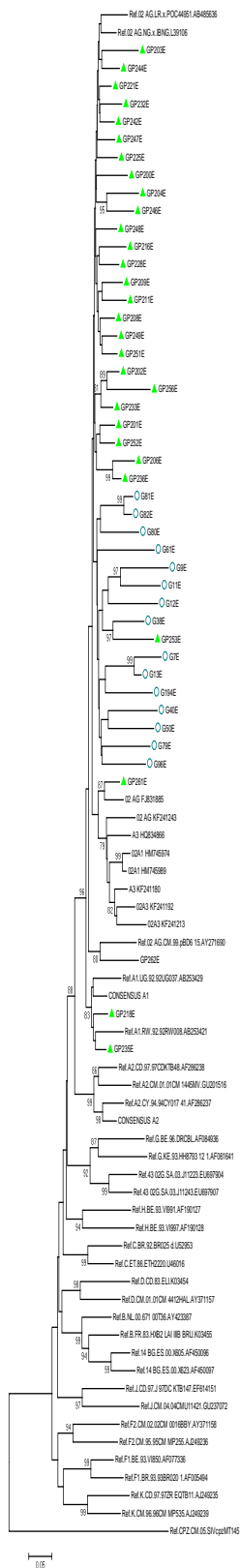


Figura 4.1. Visualização dos produtos de PCR de quatro amostras em estudo: amplificação de *env* (A); amplificação de *nef* (B); amplificação de PR (C); amplificação de RT (D). No poço 5 de cada gel foi aplicado o marcador de massa molecular 1Kb. No poço 2 em B (*nef*) e no poço 3 em C (PR), não se observou amplificação das amostras.

4.3 Caracterização Molecular das Sequências de VIH-1

4.3.1 Análise Filogenética da Região V3-V5 de *env*

Do total de 42 amostras amplificadas para a região *env* do VIH-1 foram obtidas sequências consenso para 30 amostras com uma variação de tamanho entre os 541 pb e os 653 pb. Estas sequências consenso foram utilizadas na construção da árvore filogenética juntamente com 51 sequências de referência que representam os diferentes subtipos e recombinantes do VIH-1 (Figura 4.2).



A

G

H

C

D

B

J

F2

F1

K

A

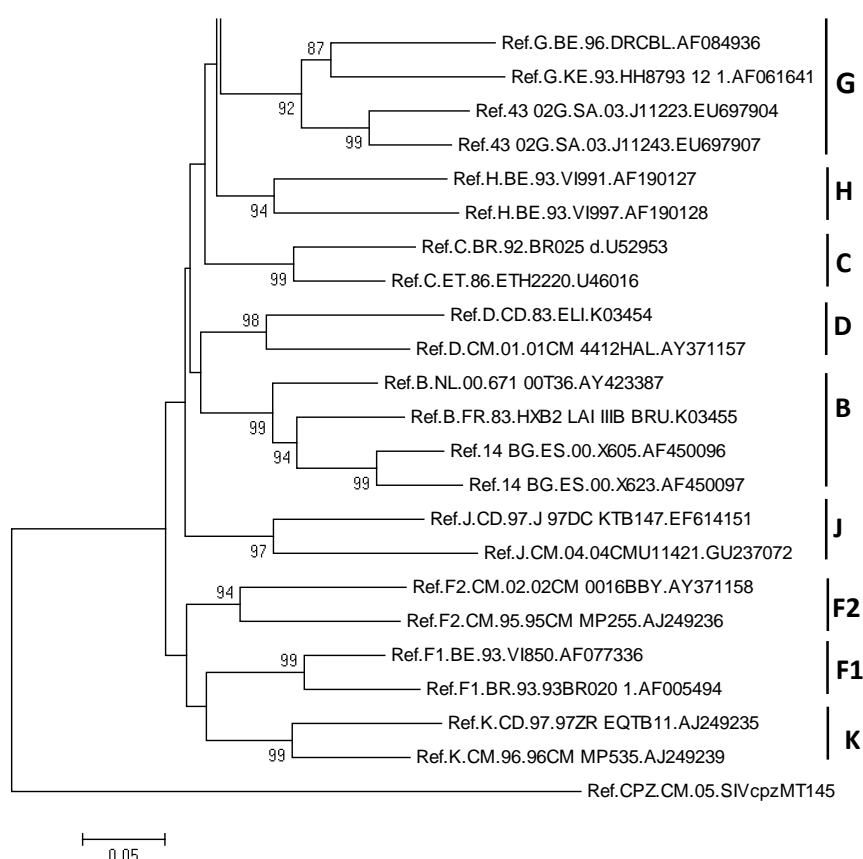


Figura 4.2. Árvore filogenética com base na análise de 588 nt da região V3-V5 do VIH-1. As sequências em estudo com prefixo GP encontram-se assinaladas com triângulos verdes (▲). As sequências com prefixo G correspondem a nativos e residentes na Guiné-Bissau e encontram-se assinalados com um círculo azul (○).

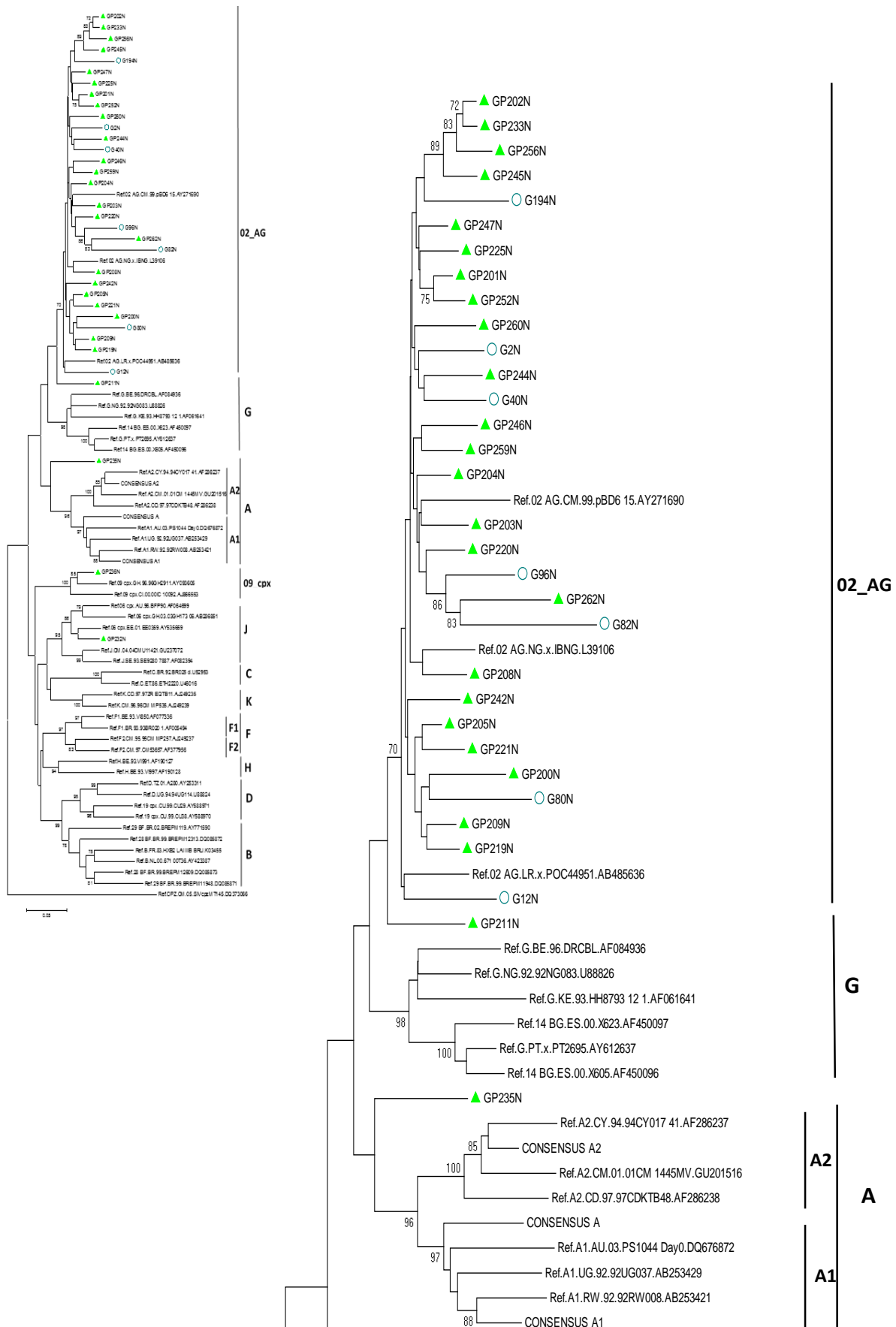
Pela topologia da árvore pode observar-se agrupamentos robustos para os diversos subtipos do VIH-1, nomeadamente os subtipos A, B, C, D, F (F1 e F2), G, H, J e K sendo suportados por elevados valores de *bootstrap*, variando de 88% a 99%.

Todas as sequências GP em estudo foram distribuídas no agrupamento de sequências classificadas em subtipo A (*bootstrap* 88%) para a região genómica analisada. Neste agrupamento observaram-se também as sequências de referência dos recombinantes CRF02_AG, CRF63_02A1 e 02A3 classificados na região *env* como subtipo A. No entanto, as sequências GP261 e GP262 agruparam com sequências de referência CRF02_AG (subtipo A em *env*), respetivamente com *bootstrap* de 87% e 80%, e as sequências GP218 e GP235 formaram um *cluster* com sequências de referência do sub-subtipo A1 suportado por um valor de *bootstrap* de 83%.

4.3.2 Análise Filogenética da Região *nef*

Das 33 amostras amplificadas para a região *nef* foram obtidas 27 sequências consenso com uma variação de tamanho entre os 621 pb e 684 pb. Para a construção da árvore filogenética foram comparadas as sequências consenso em estudo com 53 sequências de referência dos diferentes subtipos e recombinantes do VIH-1 (Figura 4.3).

Pela topologia da árvore *nef* observou-se agrupamentos de sequências com valores de *bootstrap* significativos variando entre 70% e 100% para os diversos subtipos (A, B, C, D, F, G, H, J e K) e recombinantes de VIH-1 (02_AG, 06_cpx e 09_cpx). Verificou-se que no agrupamento de sequências classificadas em CRF02_AG incluiu 23 sequências GP em estudo (*bootstrap* de 70%). Apenas quatro sequências GP em estudo foram observadas fora deste agrupamento. As sequências de referência do recombinante 09_cpx e a sequência GP236 agruparam com um elevado valor de *bootstrap* (100%). Embora a informação sobre o genoma da CRF09_cpx não esteja disponível na base de dados de *Los Alamos*, foi descrito que este recombinante está relacionado com os subtipos A ou G, ou com o recombinante CRF02_AG e a sua localização nos grupos filogenéticos depende da região em análise (McCutchan *et al*, 2004). Por outro lado, observou-se que a sequência GP232 se agrupou com sequências de referência CRF06_cpx (Anexo 1) com um valor de *bootstrap* de 86%. Por sua vez, este *cluster* de sequências observou-se no agrupamento de sequências de referências do subtipo J (*bootstrap* 95%). Na verdade, a região *nef* da CRF06_cpx foi classificada em subtipo J. Foi ainda constatado que duas sequências GP em estudo (GP211 e GP235) não foram classificadas na região genómica analisada.



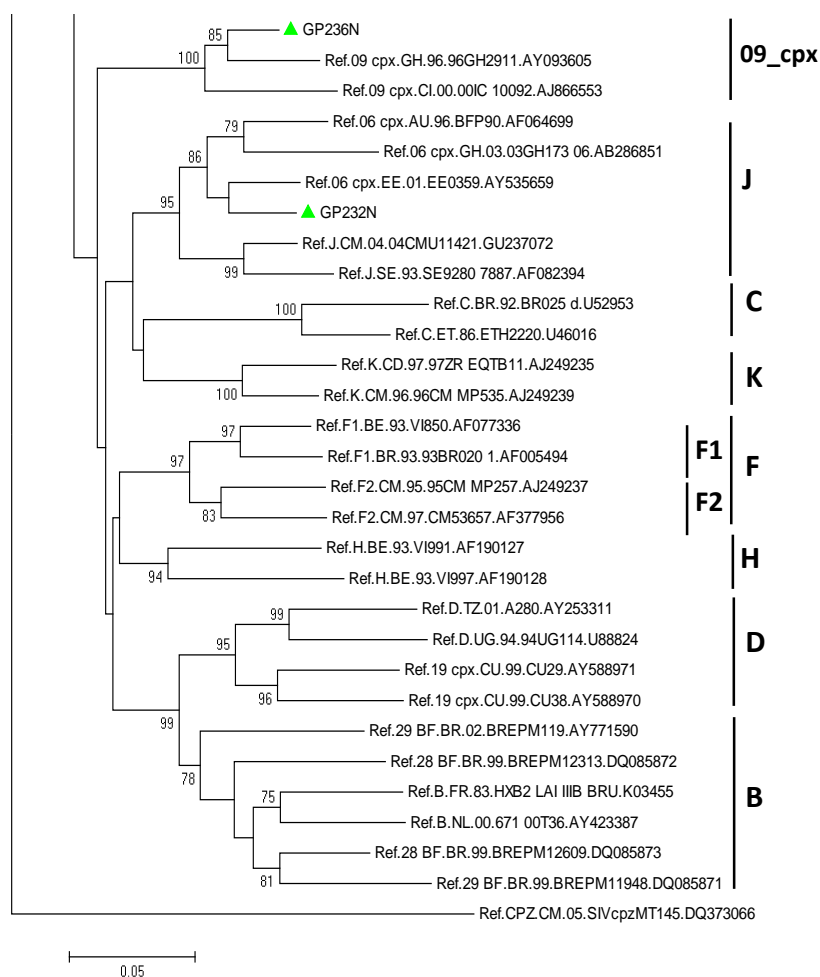


Figura 4.3. Árvore filogenética para a região *nef* do VIH-1 com base na análise de 523 nt. As sequências em estudo com prefixo GP correspondem às sequências assinaladas com triângulos verdes (▲) e as sequências com prefixo G pertencem a nativos residentes na Guiné-Bissau e encontram-se assinaladas com um círculo a azul (○).

4.3.3 Análise Filogenética da Região PR

De um total de 35 amostras amplificadas para a região PR foram obtidas 24 sequências consenso que variaram entre os 398 pb e os 474 pb. A árvore filogenética para esta região foi construída com base em 24 sequências consenso e 35 sequências de referência de diversos subtipos e recombinantes de VIH-1 (Figura 4.4). Verificou-se que as sequências formaram agrupamentos robustos e bem definidos com valores de *bootstrap* significativos, entre 74% (subtipo G), 84% (subtipo A e H) e 100% (subtipo J). Porém, não foi obtido um valor de *bootstrap* significativo para o agrupamento de referências do subtipo D (<70%).

Pela topologia da árvore observou-se que 22 sequências GP em estudo se distribuíram no agrupamento de sequências de referência CRF02_AG com valor de *bootstrap* de 86%.

Verificou-se ainda, que as restantes duas sequências em estudo (GP235 e GP254) agruparam com um valor de *bootstrap* de 73% com sequências de referência do sub-subtipo A1.

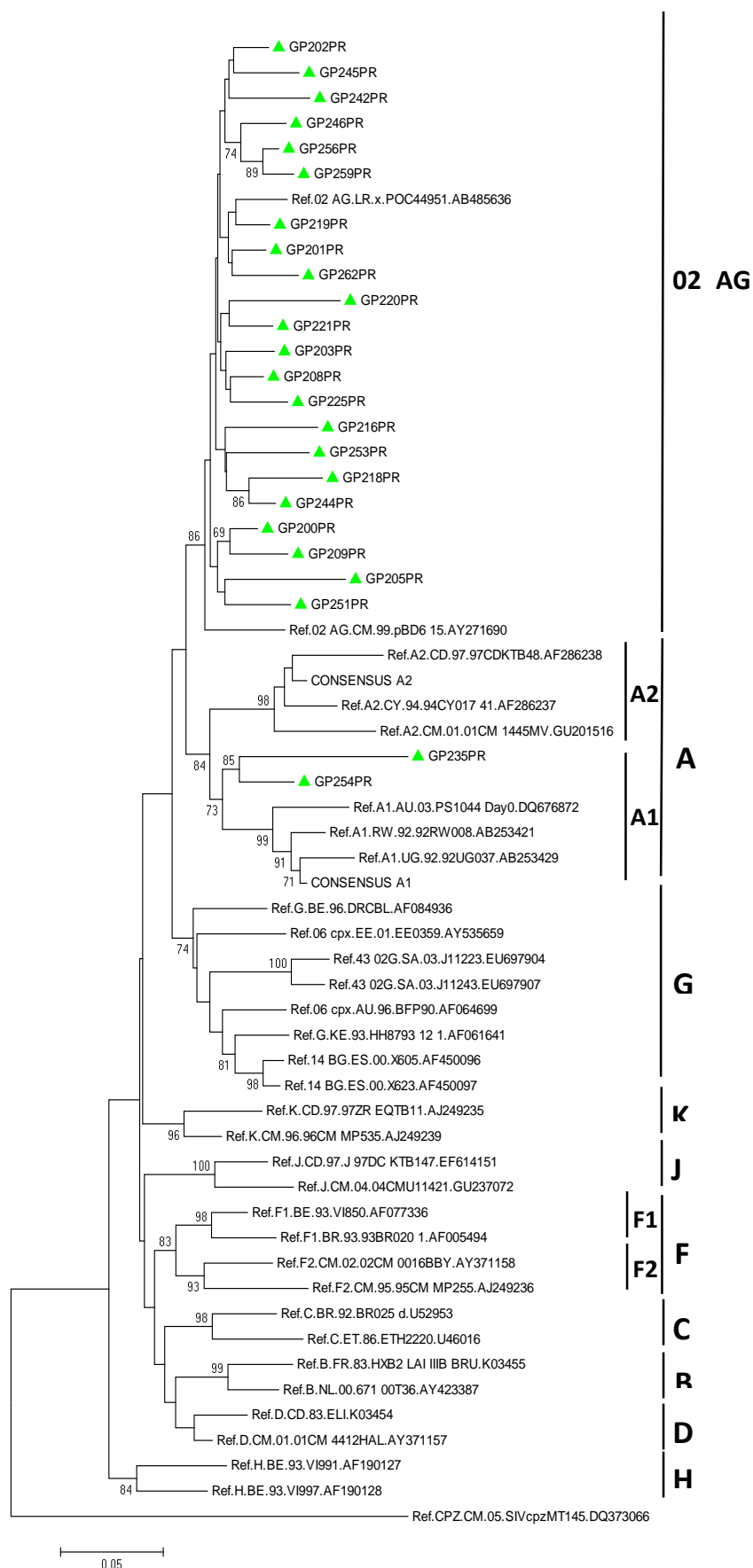


Figura 4.4. Árvore filogenética com base na análise de 450 nt da região PR do gene *pol* do VIH-1. As seqüências em estudo com prefixo GP encontram-se assinaladas com triângulos verdes (▲).

4.3.4 Análise Filogenética da Região RT

De um total de 37 amostras amplificadas para a região RT do VIH-1 foram obtidas 34 sequências de consenso variando entre os 585 pb e os 640 pb. A análise filogenética da região RT foi baseada na comparação das 34 sequências de consenso e 37 sequências de referência dos diferentes subtipos e recombinantes do VIH-1. As sequências formaram agrupamentos distintos, suportados com valores de *bootstrap* variando entre 70% (sub-subtipo A1) e 100% (subtipos C, H e J) (Figura 4.5).

As sequências de referência CRF02_AG formaram um agrupamento bem definido e robusto suportado com um valor de *bootstrap* de 76%, no qual se distribuíram 30 sequências GP em estudo.

Verificou-se que as sequências GP235 e GP254 agruparam com valor de *bootstrap* de 99%, no entanto, foram incluídas no agrupamento de sequências de referência do sub-subtipo A1 com um valor de *bootstrap* de 70%. As restantes 2 sequências GP em estudo (GP206 e GP236) formaram um agrupamento isolado com *bootstrap* de 100%, não sendo possível classificá-las pela análise filogenética realizada na região RT do VIH-1.

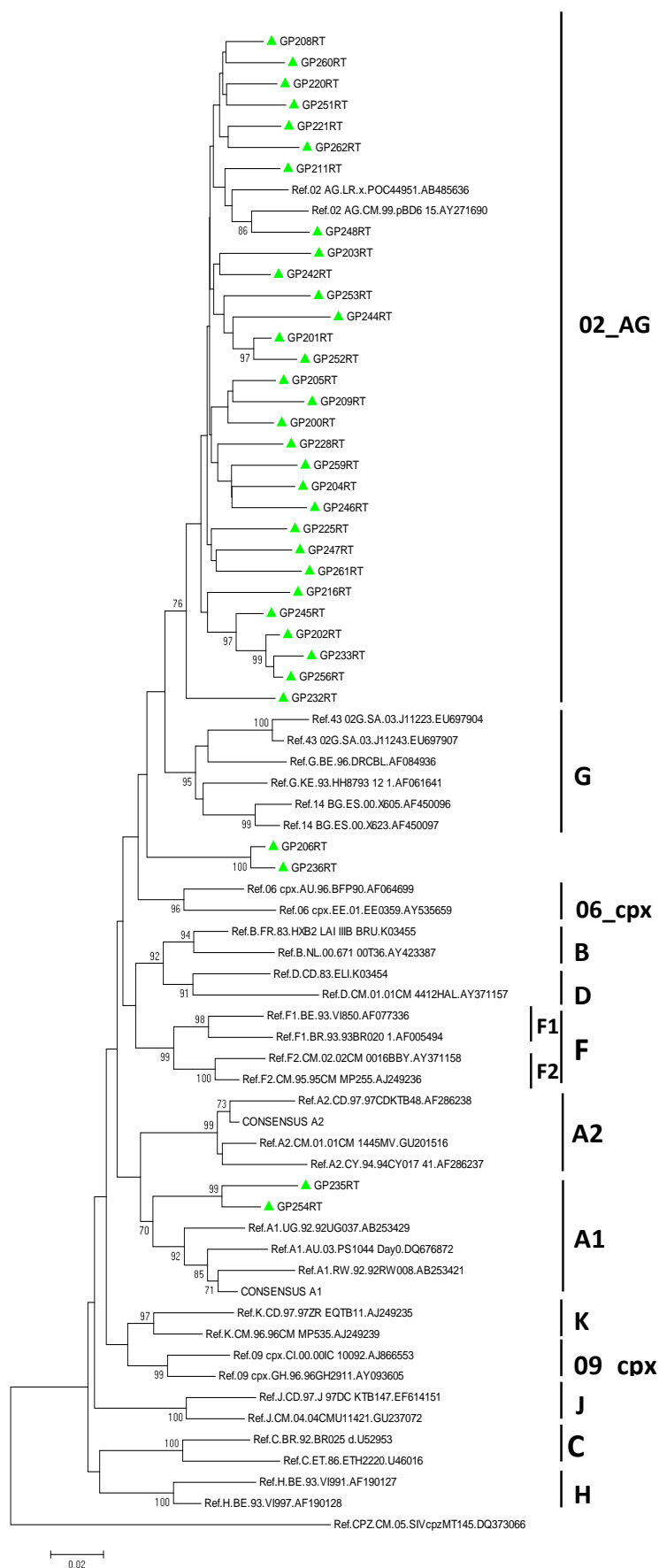


Figura 4.5. Árvore filogenética com base na análise de 624 nt da região RT do gene *pol* do VIH-1. Os triângulos verdes (▲) assinalam as sequências GP em estudo.

4.3.5 Análise Conjunta das Regiões V3-V5, *nef*, PR e RT

Para a classificação das amostras VIH-1 procedeu-se a uma análise molecular conjunta dos resultados da estimativa filogenética obtida para as sequências *env*, *nef*, PR e RT do VIH-1 (Tabela 4.1). Verificou-se que em 13 (35%) amostras foi obtida uma classificação do VIH-1 a partir de quatro regiões genómicas, e em igual número (n=13; 35%) de amostras, a partir de três regiões genómicas. A classificação do VIH-1 em 9 (25%) amostras foi baseada em duas regiões genómicas e nas restantes 2 (5%) amostras a classificação foi limitada apenas à região *env* do VIH-1 (GP206 e GP249).

Do total, das 37 amostras GP em estudo, verificou-se que 30 (82%) foram classificadas em CRF02_AG e 2 (5%) em sub-subtipo A1. Observou-se ainda que a classificação de 2 (5%) sequências em subtipo A foi limitada à região *env*, pelo que poderá não corresponder a uma classificação correta do VIH-1. De facto, várias amostras correspondem à infeção CRF02_AG tendo por base a análise de três regiões genómicas do VIH (*nef*, PR e RT) foram classificadas em subtipo A na região *env*. De notar que nesta região genómica a CRF02_AG agrupa com sequências de referência do subtipo A.

Por outro lado, a classificação discordante obtida em duas ou três regiões genómicas analisadas para as restantes 3 (8%) amostras indicou que podem ser formas recombinantes. Assim, no caso GP218 poderá ter ocorrido uma recombinação entre a CRF02_AG e o sub-subtipo A1, no caso GP232 uma recombinação entre a CRF02_AG, subtipo A e a CRF06_cpx e no caso GP236 entre o subtipo A e a CRF09_cpx. Para confirmar a classificação deste vírus seria necessário a análise do genoma completo do VIH-1.

Tabela 4.1. Classificação filogenética das sequências GP em estudo, com base nos resultados obtidos nas quatro regiões genômicas do VIH-1.

Classificação Filogenética das Sequências				
Amostra	Região V3-V5	Gene <i>nef</i>	Região PR	Região RT
GP200	A	02_AG	02_AG	02_AG
GP201	A	02_AG	02_AG	02_AG
GP202	A	02_AG	02_AG	02_AG
GP203	A	02_AG	02_AG	02_AG
GP204	A	02_AG	n.s	02_AG
GP205	n.s	02_AG	02_AG	02_AG
GP206	A	n.a	n.s	n.c
GP208	A	02_AG	02_AG	02_AG
GP209	A	02_AG	02_AG	02_AG
GP211	A	n.c	n.s	02_AG
GP216	A	n.a	02_AG	02_AG
GP218	A ¹	n.s	02_AG	n.s
GP219	n.s	02_AG	02_AG	n.s
GP220	n.s	02_AG	02_AG	02_AG
GP221	A	02_AG	02_AG	02_AG
GP225	A	02_AG	02_AG	02_AG
GP228	A	n.a	n.s	02_AG
GP232	A	06_cpx ³	n.a	02_AG
GP233	A	02_AG	n.s	02_AG
GP235	A ¹	n.c	A ¹	A ¹
GP236	A	09_cpx	n.s	n.c
GP242	A	02_AG	02_AG	02_AG
GP244	A	02_AG	02_AG	02_AG
GP245	n.s	02_AG	02_AG	02_AG
GP246	A	02_AG	02_AG	02_AG
GP247	A	02_AG	n.s	02_AG
GP248	A	n.a	n.a	02_AG
GP249	A	n.a	n.a	n.a
GP251	A	n.a	02_AG	02_AG
GP252	A	02_AG	n.s	02_AG
GP253	A	n.s	02_AG	02_AG
GP254	n.s	n.s	A ¹	A ¹
GP256	A	02_AG	02_AG	02_AG
GP259	n.s	02_AG	02_AG	02_AG
GP260	n.s	02_AG	n.a	02_AG
GP261	A ²	n.a	n.s	02_AG
GP262	A ²	02_AG	02_AG	02_AG

¹ Na árvore filogenética agrupou com sequências A1; ² Na árvore filogenética agrupou com sequências CRF02_AG; ³ Na árvore filogenética agrupou com sequências J; n.a – Não amplificada; n.s – não sequenciada; n.c – Não classificada por filogenia.

4.3.6. Distâncias nucleotídicas entre sequências da região V3-V5 de *env* e gene *nef*

As distâncias nucleotídicas foram calculadas no grupo de sequências GP em estudo e no grupo de sequências de referências G obtidas de indivíduos naturais e a residir na Guiné-Bissau, para as duas regiões genómicas (*nef* e *env*). Na tabela 4.2 encontram-se apresentados os valores percentuais médios das distâncias nucleotídicas obtidos por região genómica analisada e nos dois grupos de sequências analisados.

Tabela 4.2. Cálculo das distâncias nucleotídicas intra e inter grupo para as sequências V3-V5 e *nef* do VIH-1.

Regiões do VIH-1	Distâncias nucleotídicas entre sequências (%)	
	Sequências do grupo G	Sequências do grupo GP
V3-V5	22,4	13,0
<i>nef</i>	12,0	9,2

Com o intuito de comparar os resultados das distâncias nucleotídicas intra e intergrupo recorreu-se a uma sequência de referência CRF02_AG (L39106) utilizada para o cálculo dos valores em ambos os grupos de sequências (G e GP). Assim, para a região V3-V5 do VIH-1 obteve-se distâncias genéticas de 18,4% e 10,9%, respetivamente, para as sequências do grupo G e para as sequências do grupo GP. Por sua vez, para a região *nef* obteve-se distâncias genéticas de 10,0% e 9,0%, respetivamente para as sequências do grupo G e para as sequências do grupo GP.

Numa análise geral das distâncias nucleotídicas observou-se que tanto para a região *env*, como para a região *nef* os valores médios das distâncias genéticas foram superiores nas sequências G de indivíduos naturais e residentes na Guiné comparativamente às sequências de Guineenses residentes em Portugal. Por outro lado, comparando as duas regiões genómicas, verificou-se que os valores obtidos para *env* foram sempre superiores aos valores de distância nucleotídica encontrada para *nef*, independentemente dos grupos de sequências analisadas.

4.4 Análise das Sequências Aminoacídicas de VIH-1

4.4.1 Região V3-V5

A análise aminoacídica da região V3 (35 resíduos) foi realizada com base num alinhamento das 30 sequências em estudo (Figura 4.6).

Analisando com maior detalhe a constituição aminoacídica desta região, verificou-se que com a exceção da sequência GP211 que apresentou a deleção de um aminoácido na posição 24, não foram encontradas outras deleções ou inserções nas restantes sequências analisadas. O principal motivo estrutural, GPG, encontra-se totalmente conservado em todas as sequências estudadas. O mesmo se observou no motivo responsável pela conformação em *loop* da região V3, composto pelas duas cisteínas, C₁ e C₃₅ que se apresenta 100% conservadas.

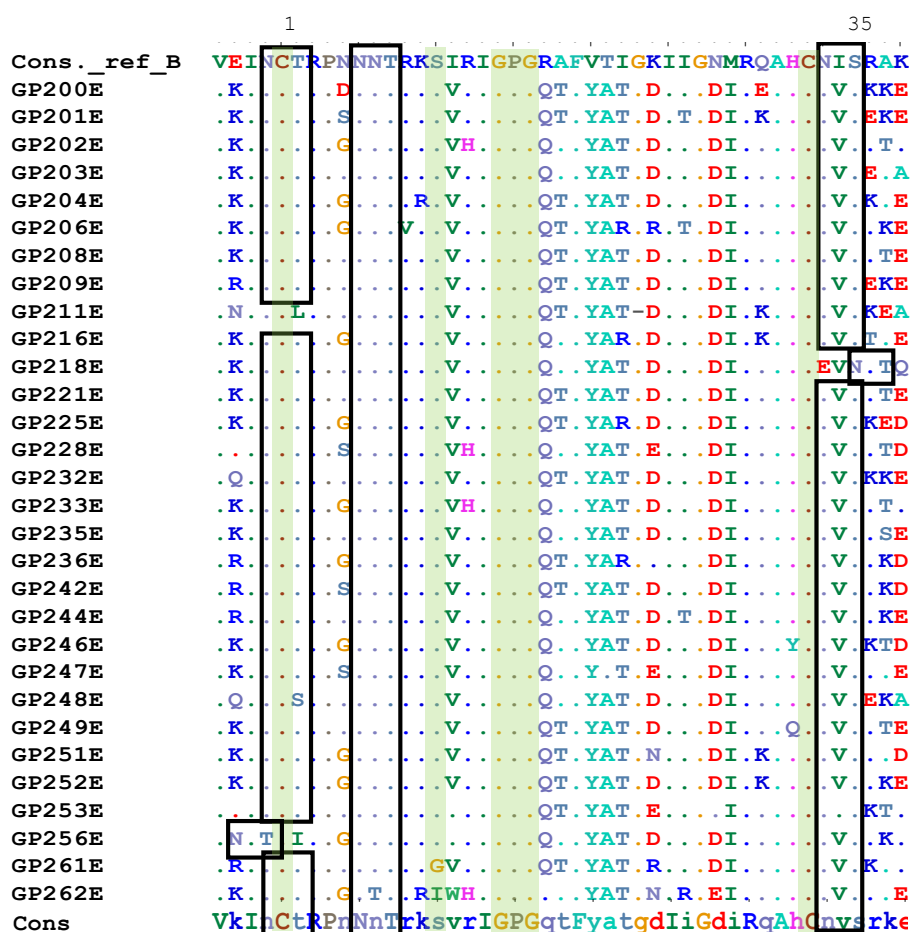


Figura 4.6. Alinhamento parcial da região V3-V5 de *env* do VIH-1. As sequências em estudo foram comparadas com uma sequência de referência do subtipo B. Os pontos representam a semelhança entre as sequências e os traços correspondem a *gaps* resultantes do alinhamento. Os retângulos assinalam os locais de N-glicosilação e a verde os motivos estruturais analisados.

Na análise global das sequências aminoacídicas da região V3- V5, verificou-se que foram encontrados alguns polimorfismos. De uma forma geral, observou-se para determinadas posições no alinhamento uma frequência elevada de resíduos comuns entre as sequências GP em estudo. No entanto, esses resíduos são diferentes dos identificados na sequência de referência do subtipo B (Tabela 4.3). De referir que as sequências GP em estudo foram classificadas em subtipo A na região *env* do VIH-1.

Tabela 4.3. Polimorfismos mais frequentes (%) encontrados na sequência aminoacídica V3-V5 do VIH-1.

Posição do Resíduo	Resíduo na Sequência de Referência B	Resíduo mais frequente nas Sequências em estudo (%)
12	I	V (90%)
18	R	Q (97%)
21	V	Y (100%)
22	T	A (97%)
23	I	T (87%)
25	K	D (83%)
29	N	D (93%)
30	M	I (100%)

Foi observada uma elevada conservação dos locais de N-glicosilação (NxS/T), nas 30 sequências GP em estudo. Com exceção de sequência GP211 em que se identificou apenas dois locais de N-glicosilação, as restantes sequências mantiveram os três locais de N-glicosilação.

a) Previsão dos Co-recetores utilizados

Adicionalmente à análise dos locais de N-glicosilação na região V3 também foi realizada a previsão do uso dos co-recetores CCR5 ou CXCR4 no grupo de amostras VIH-1 estudadas (Tabela 4.4).

Tabela 4.4. Previsão do co-recetor utilizado nas amostras GP em estudo

Amostras VIH-1 estudadas (n=30)			
Sequência	Co-recetor	Sequência	Co-recetor
GP200E	CCR5	GP233E	CCR5
GP201E	CCR5	GP235E	CCR5
GP202E	CCR5	GP236E	CXCR4
GP203E	CCR5	GP242E	CCR5
GP204E	CCR5	GP244E	CCR5
GP206E	CXCR4	GP246E	CCR5
GP208E	CCR5	GP247E	CCR5
GP209E	CCR5	GP248E	CCR5
GP211E	CCR5	GP249E	CCR5
GP216E	CCR5	GP251E	CCR5
GP218E	CCR5	GP252E	CCR5
GP221E	CCR5	GP253E	CCR5
GP225E	CCR5	GP256E	CCR5
GP228E	CCR5	GP261E	CCR5
GP232E	CCR5	GP262E	CCR5

Em 28 (93%) dos casos analisados foi prevista a utilização do co-recetor CCR5 e em apenas 2 (7%) casos a utilização do co-recetor CXCR4 (GP206 e GP236).

4.4.2. Proteína Nef

A análise da região Nef foi realizada através do alinhamento parcial de 27 sequências, contendo os principais motivos estruturais associados a funções da proteína (Figura 4.7).

Assim, da extremidade 5' para 3' de Nef verificou-se que:

- O sinal de meristilação (MGxxS₁) encontra-se 100% conservado.
- O resíduo de metionina (M₂₀) encontra-se alterado para o resíduo de isoleucina em 20 (74,1%) sequências analisadas.
- O motivo RRAE₂₁, representado na sequência de referência do subtipo B por QRAE₂₁ foi observado nas formas R/KQTP em 23 (85,2%) sequências GP em estudo.
- O motivo W₅₇ apresentou-se 100% conservado.

- O motivo EEEE₆₂ apenas foi observado conservado em quatro sequências (GP205, GP208, GP225 e GP262), enquanto que nas restantes sequências foram observadas alterações de E para D/G/Q no 2º, 3º e 4º resíduos.
- O motivo da tretraprolina PxxPxxPxRP apresentou-se 100% conservado.
- O motivo RPMTYK₇₇ apresentou-se alterado em 8 (29,6%) sequências com substituição do resíduo de tirosina (Y) pelos resíduos de fenilalanina (F) ou serina (S).
- O motivo RR₁₀₅ foi observado em 24 (88,9%) sequências em estudo na forma KR₁₀₅ matendo a forma da sequência de referência do subtipo B. Apenas 2 sequências em estudo apresentaram a forma RR₁₀₅ e uma sequência a forma KK₁₀₅.
- O domínio FPD₁₂₁ encontra-se 100% conservado.
- O motivo de prolina PxxP₁₄₇ foi observado conservado, apenas com a exceção da sequência GP235, que se observou na forma PxxK₁₄₇.
- O motivo diacídico EE₁₅₄ foi observado conservado em 21 (77,8%) sequências, no entanto verificou-se que 5 sequências apresentaram a forma do motivo igual à forma da sequência de referência do subtipo B (EK₁₅₄) e uma sequência o motivo na forma KE₁₅₄.
- No domínio ExxxLL, os dois resíduos de leucina observaram-se 100% conservados.
- O motivo terminal DDPxxE, apresentou-se alterado em quase todas as sequências observando-se na forma EDExxE na maioria das sequências em estudo.

Tabela 4.5. Polimorfismos mais frequentes encontrados na sequência aminoacídica de Nef do VIH-1.

Posição do Resíduo	Resíduo na Sequência de Referência B	Resíduo mais frequente nas Sequências em estudo (%)
20	M	I (74%)
21	Q	R (78%)
22	R	Q (85%)
23	A	T (100%)
24	E	P (96%)
148	V	M (81%)
155	K	E (81%)
163	C	S (85%)
174	D	E (93%)
176	P	E (70%)
177	E	D (81%)
178	K	R (81%)

Nas sequências GP em estudo foi encontrada uma elevada frequência de resíduos comuns entre as sequências, contudo, diferentes dos observados na sequência de referência do subtipo B, sugerindo tratar-se de polimorfismos associados a sequências de subtipos não-B (Tabela 4.5). Na verdade, a maioria das sequências GP em estudo foram classificadas em CRF02_AG.

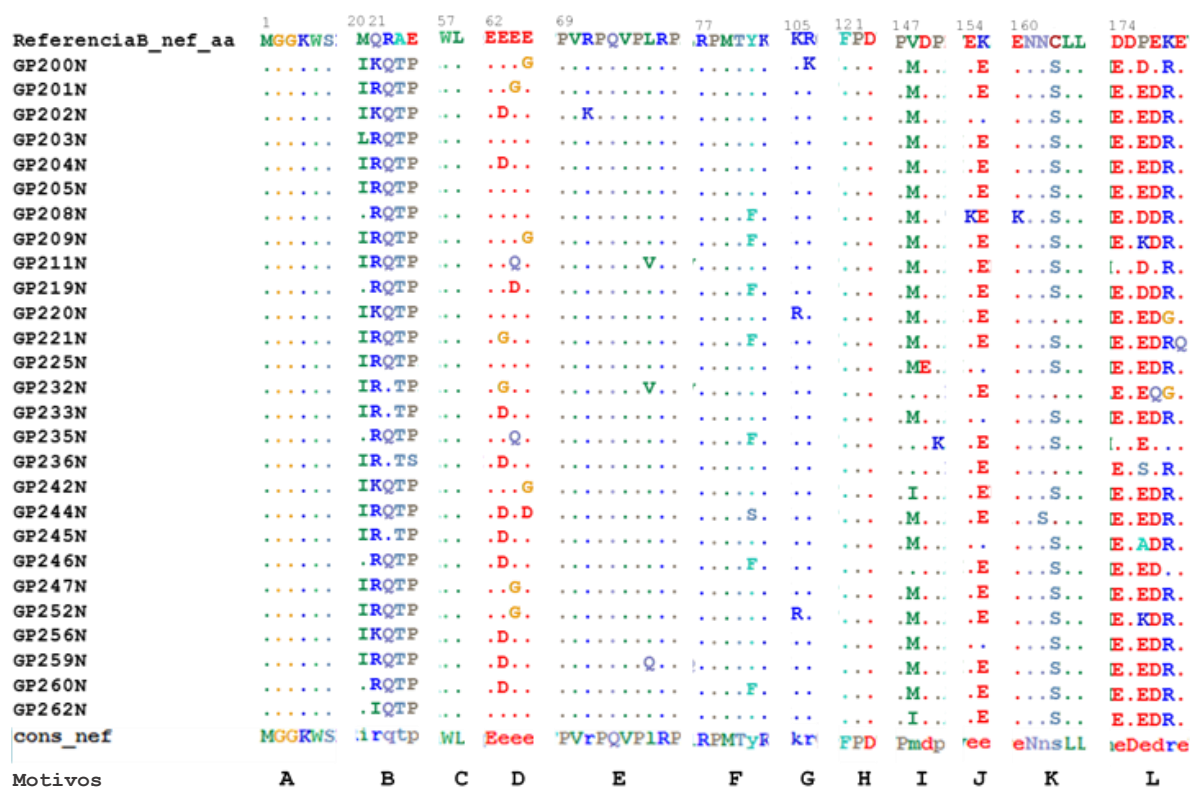


Figura 4.7. Alinhamento parcial de Nef As sequências GP em estudo e a sequência de referência B (em cima e em baixo). As letras maiúsculas e minúsculas indicam respectivamente o consenso de Nef total e parcial. Os pontos representam a semelhança entre as posições das sequências. Os motivos estruturais e funcionais de Nef estão designados respectivamente pelas letras maiúsculas de A a L (em baixo). A- Sinal de Meristilação; B- M₂₀ e a Hélice α; C- Motivo WL₅₇; D- Motivo Acídico; E- Motivo da Tetrapirolina; F- Motivo RPMTYK; G- Motivo RR₁₀₅; H- Motivo FPD; I- Motivo da Prolina; J- Motivo EE₁₅₄; K- Motivo de Dileucina; L- Motivo DDPxxE.

4.4.3. Região PR

O alinhamento das 24 sequências em estudo da região PR revela a conservação dos motivos DTGAD₂₅ e IGGIGGF₄₇ e do resíduo I₈₄ (Figura 4.8). A análise do alinhamento sugere a existência de algumas posições polimórficas, nomeadamente as posições 3, 20, 36, 41, 69 e 89, que apresentam um resíduo alterado comparativamente à sequência de referência do subtipo B, mas que se mantém alterado na maioria das sequências GP em estudo.

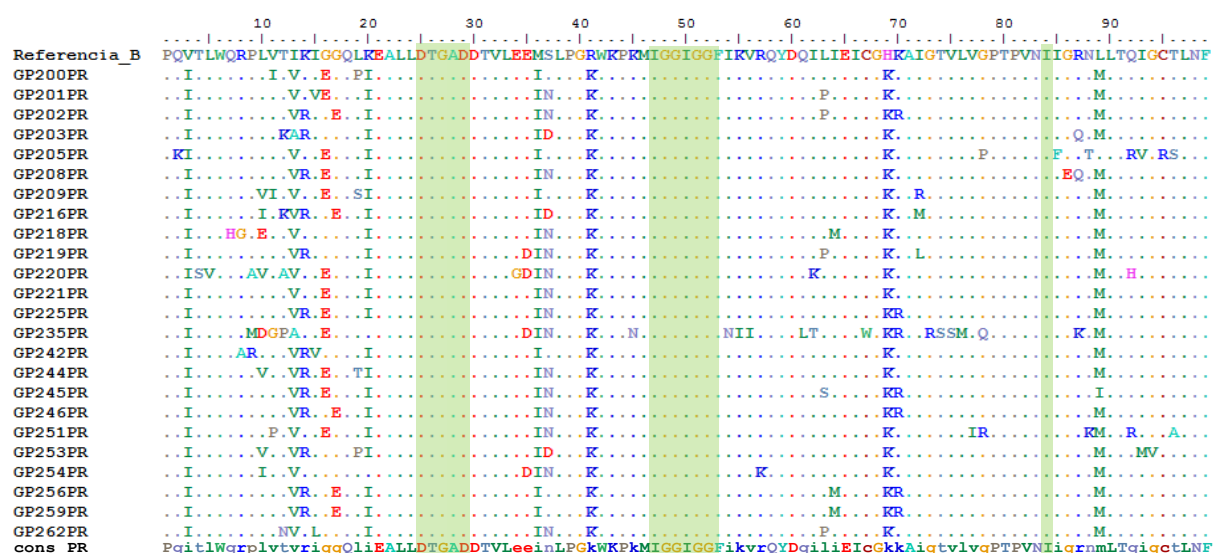


Figura 4.8. Alinhamento da Região PR contendo as sequências em estudo e de referência do subtipo B. As letras maiúsculas e minúsculas (em cima e em baixo) representam o consenso total e parcial e os pontos indicam a semelhança entre as sequências. O sombreado a verde assinala os motivos estruturais analisadas.

No total de 24 sequências analisadas para a região PR, observou-se três mutações de resistência em quatro sequências GP em estudo (Tabela 4.6). As mutações identificadas foram a V11I nas sequências GP200 e GP216, a T74P na GP235 e a N88T na GP205. Adicionalmente, em 22 (92%) sequências em estudo foi identificada a mutação K20I e em 21 (88%) sequências a mutação L89M, no entanto estas alterações foram descritas como polimorfismos de subtipos não-B, não conferindo por si só resistência aos PI.

Tabela 4.6. Frequência das mutações de resistência aos PI nas sequências GP em estudo

Mutações de resistência aos PI	Número de mutações de resistência (n=24)	Sequências GP
V11I	2	GP200; GP216
T74P	1	GP235
N88T	1	GP205

4.4.4 Região RT

Na análise do alinhamento parcial da região RT de 34 sequências GP em estudo observou-se o motivo YxDD 100% conservado assim como os resíduos K₆₅, R₇₂, Y₁₁₅, Q₁₅₁ e os resíduos da tríade catalítica, D₁₁₀, D₁₈₅ e D₁₈₆ (Figura 4.9).

	65	72	110	115	151	183
Referencia_B	K	R	D	Y	Q	YmDD
GP200RT	*	*	*	*	*
GP201RT	*	*	*	*	*
GP202RT	*	*	*	*	*
GP203RT	*	*	*	*	*
GP204RT	*	*	*	*	*
GP205RT	*	*	*	*	*
GP206RT	*	*	*	*	*
GP208RT	*	*	*	*	*
GP209RT	*	*	*	*	*
GP211RT	*	*	*	*	*
GP216RT	*	*	*	*	*
GP220RT	*	*	*	*	*
GP221RT	*	*	*	*	*
GP225RT	*	*	*	*	*
GP228RT	*	*	*	*	*
GP232RT	*	*	*	*	*
GP233RT	*	*	*	*	*
GP325RT	*	*	*	*	*
GP236RT	*	*	*	*	*
GP242RT	*	*	*	*	*
GP244RT	*	*	*	*	*
GP245RT	*	*	*	*	*
GP246RT	*	*	*	*	*
GP247RT	*	*	*	*	*
GP248RT	*	*	*	*	*
GP251RT	*	*	*	*	*
GP252RT	*	*	*	*	*
GP253RT	*	*	*	*	*
GP254RT	*	*	*	*	*
GP256RT	*	*	*	*	*
GP259RT	*	*	*	*	*
GP260RT	*	*	*	*	*
GP261RT	*	*	*	*	*
GP262RT	*	*	*	*	*	..V..
cons_RT	K	R	D	Y	Q	YmDD

Figura 4.9. Alinhamento parcial da região RT das sequências GP em estudo e a sequências referência do subtipo B. As letras maiúsculas e minúsculas (em cima e em baixo) indicam, respectivamente, o consenso total e parcial e os pontos correspondem à semelhança entre as sequências.

No total de 34 sequências analisadas da região RT, verificou-se a presença de mutações de resistências aos NNRTI em 4 (12%) sequências GP em estudo. A mutação V90I na sequência GP208, a mutação A98G nas sequências e GP221 e GP262, a mutação K103N na sequência

GP262 e a mutação V106I na sequência GP252. Por outro lado, foi observada uma única mutação de resistência aos NRTI, a M184V identificada na sequência GP262. Deste modo, verificou-se que a sequência GP262 apresentou um total de 3 mutações de resistência aos RTI (Tabela 4.7).

Tabela 4.7. Frequência das mutações de resistência aos NNRTI e NRTI nas sequências GP em estudo

Mutações de resistência aos NNRTI	Número de mutações de resistência (n=34)	Ref ^{as} das sequências GP	Mutações de resistência aos NRTI	Número de mutações de resistência (n=34)	Ref ^{as} das Sequências GP
V90I	1	GP208	M184V	1	GP262
A98G	2	GP221; GP262			
K103N	1	GP262			
V106I	1	GP252			

5. Discussão

O Vírus da Imunodeficiência Humana caracteriza-se por uma elevada diversidade e contínua capacidade de mutação e recombinação genética resultando na produção de um número crescente de variantes virais ao longo do tempo. Este facto tem vindo a causar problemas no combate à infeção VIH/SIDA, pelo ao aparecimento e seleção de variantes com mutações de resistência aos antirretrovirais.

O presente estudo pretendeu determinar a diversidade genética do VIH-1 num grupo de indivíduos Guineenses residentes em Portugal e comparar os resultados com um estudo desenvolvido no laboratório em nativos residentes na Guiné-Bissau. Para comparar o padrão epidemiológico molecular do VIH-1, as amostras de Guineenses residentes em Portugal foram caracterizadas nas regiões genómicas *env* e *nef*, seguindo o modelo de estudo desenvolvido anteriormente em amostras de nativos residentes na Guiné-Bissau. Adicionalmente, as amostras foram analisadas na região *pol* de VIH-1 para classificação do vírus e identificação de mutações de resistência associadas aos inibidores da PR e RT.

No estudo desenvolvido foi analisada uma amostra de conveniência. As amostras foram colhidas a 42 mulheres jovens (média de 30,9 anos de idade) que tiveram um filho entre 2 a 30 dias antes da data de colheita. A maioria das mulheres (n=38; 90%) fez prevenção da transmissão mãe-filho do VIH-1 seguindo os regimes terapêuticos recomendados (inibidores da PR e RT). Observou-se que no grupo estudado apenas quatro mulheres (GP232, GP245, GP249 e GP256) não cumpriram a prevenção, e uma delas, transmitiu o VIH-1 ao filho (GP256).

Para a classificação das amostras procedeu-se à amplificação de ADN pela técnica de *nested*-PCR para diferentes regiões genómicas do VIH-1. Obteve-se um sucesso de 100% para a amplificação de *env*, 86% para RT, 83% para PR e 79% para *nef*. O insucesso obtido na amplificação de algumas das amostras pode ser explicado pela degradação do ADN, devido aos vários ciclos de congelamentos e descongelamentos das amostras.

Das amostras amplificadas e sequenciadas foi possível construir as respetivas sequências de consenso em 92% para a região RT (34/37), 82% para o gene *nef* (27/33), 70% para *env* (30/42) e 69% para PR (24/35). Nas sequências de consenso obtidas para cada uma das regiões genómicas do VIH-1 foram observadas variações no tamanho do fragmento sequenciado, que poderão dever-se ao baixo grau de definição dos picos obtidos na fase final da sequenciação

do fragmento. Em alguns cromatogramas constatou-se a presença de mistura de sequências virais na amostra.

Para classificação do VIH-1 foi realizada a análise filogenética das regiões *env*, *nef*, PR e RT através da construção das respetivas árvores filogenéticas. Na construção das árvores foi utilizado o método de *Neighbour-Joining* (Saitou & Nei, 1987), frequentemente utilizado pela rapidez, fiabilidade e eficiência na obtenção de resultados para a classificação de sequências VIH-1.

A árvore filogenética correspondente à região *env* do VIH-1 permitiu classificar a totalidade das sequências GP em estudo em subtipo A nesta região genómica. De salientar que no agrupamento A se observou as sequências de referência CRF02_AG (subtipo A em *env*). Observou-se ainda que duas (7%) sequências em estudo (GP218 e GP235) agruparam com sequências de referências do sub-subtipo A1.

De uma forma geral verificou-se pela topologia da árvore de *env* uma separação entre a maioria das sequências GP (Guineenses residentes em Portugal) e das sequências de referência G (Guineenses residentes na Guiné-Bissau), ou seja, embora todas as sequências fossem incluídas no agrupamento do subtipo A, observou-se uma maior semelhança entre as sequências pertencentes a cada um dos dois grupos (GP e G).

A análise da topologia da árvore da região *nef* do VIH-1 não permitiu classificar duas (7%) sequências GP em estudo. Das restantes 25 (93%) sequências, 23 (85%) foram classificadas em CRF02_AG, uma (4%) classificada como CRF09_cpx e outra (4%) classificada em CRF06_cpx.

A sequência GP232 foi classificada na região *nef* em CRF06_cpx. No entanto, a região *nef* deste recombinante é subtipo J e agrupou como seria de esperar, com sequências de referência do subtipo J.

Por outro lado, a sequência GP236, encontra-se num *cluster* com as sequências de referência CRF09_cpx (100% de *bootstrap*). No entanto, este recombinante descrito por McCutchan *et al* (2004), pode agrupar com sequências de referência dos subtipos A ou G ou pode ser observado com sequências de referência CRF02_AG dependendo da região genómica do VIH-1 analisada. No entanto, os agrupamentos de sequências dos subtipos A, G e CRF02_AG

observados na árvore filogenética de *nef* do VIH-1 não incluem as sequências de referência da CRF09_cpx.

Contrariamente ao observado na árvore filogenética da região *env*, na região *nef* as sequências de referência G (nativos residentes na Guiné-Bissau) parecem estar aleatoriamente distribuídas no agrupamento da CRF02_AG, juntamente com as sequências GP em estudo.

Verificou-se que a totalidade das sequências GP em estudo foram classificadas na região PR, nomeadamente, 22 (92%) em CRF02_AG e 2 (8%) em sub-subtipo A1 (GP235 e GP254).

A análise da topologia da árvore da região RT (n=34) não permitiu classificar duas (6%) das sequências GP em estudo (GP206 e GP236). No entanto, de igual modo as restantes 30 (88%) sequências, foram classificadas em CRF02_AG e duas (6%) sequências em sub-subtipo A1 (GP235 e GP254).

Assim, numa análise conjunta de resultados de classificação do VIH-1 para 37 amostras em estudo, verificou-se que a maioria das sequências apresentou classificação concordante relativamente às regiões genómicas analisadas, sendo que 82% (n=30) das sequências GP em estudo foram classificadas como potenciais CRF02_AG e 5% (n=2) classificadas em sub-subtipo A1 (GP235 e GP254). No entanto, os resultados da análise filogenética para as restantes cinco amostras necessita de mais alguma reflexão. Assim, verificou-se que as sequências GP206 e GP249 foram classificadas de subtipo A. Porém, este resultado foi baseado na análise de uma pequena porção do genoma do VIH-1 (região *env*), o que pode conduzir a uma classificação incorreta do vírus. De facto, outras sequências GP em estudo classificadas, em subtipo A na região *env* do VIH-1, revelaram ser CRF02_AG quando a análise foi alargada a outras regiões genómicas. Por outro lado, uma classificação discordante das sequências nas diferentes regiões genómicas de VIH-1 (GP218, GP232 e GP236) sugere que o vírus é uma forma recombinante com um potencial genoma mosaico.

Na classificação da sequência GP218, constatou-se uma provável recombinação entre a CRF02_AG e o sub-subtipo A1 (classificação distinta na região PR e *env*). A sequência GP232, revelou também uma classificação distinta nas três regiões analisadas: CRF02_AG em RT,

subtipo A em *env* e 06_cpx em *nef* (subtipo J em *nef*). A classificação da sequência GP236 nas regiões *env* e *nef* sugeriu recombinação entre o subtipo A e a CRF09_cpx.

Neste estudo foi observada uma elevada prevalência de CRF02_AG, estando este resultado de acordo com o descrito sobre a epidemiologia molecular da infeção VIH-1 na Guiné-Bissau (Andersson *et al*, 1999; Esbjornsson *et al*, 2011). De salientar que a CRF02_AG também se encontra em circulação na população portuguesa (Abecasis *et al*, 2013). A elevada frequência e distribuição da CRF02_AG deve-se à facilidade de transmissão do vírus como consequência da elevada virémia e alta taxa de replicação viral, como descrito por Njai *et al* (2006). Por outro lado, na população em estudo identificaram-se também três variantes com genoma mosaico que podem ser potencialmente URF (CRF02_AG/A1; CRF02_AG/A/J; A/09_cpx), atendendo que não se encontram identificadas na base de dados formas recombinadas semelhantes às obtidas a partir destas três amostras.

Os resultados obtidos na análise conjunta das quatro regiões analisadas neste estudo confirmam a existência, em Portugal, de um perfil molecular distinto dos restantes países europeus, devido à elevada frequência de subtipos não-B (Esteves *et al*, 2002; Palma *et al*, 2007). O facto deste estudo ser constituído exclusivamente por mulheres que adquiriram a infeção através de transmissão heterossexual, corrobora também os dados obtidos, uma vez que estas características (sexo e via de transmissão) parecem corresponder a subtipos não-B (Abecasis *et al*, 2013). Curiosamente, e apesar da sua origem e elevada prevalência no nosso país não foram identificados casos de infeção pela CRF14_BG. A razão para este facto pode ser devido a este recombinante estar associado a grupos populacionais de UDI (Niculescu *et al*, 2015). Contudo, esta forma recombinante não é descrita na Guiné-Bissau.

Para analisar o grau de divergência entre as sequências V3-V5 e *nef* em estudo, foram calculadas as distâncias genéticas entre o grupo de sequências GP estudadas e o grupo de sequências de referência G, derivadas de amostras de indivíduos naturais e residentes na Guiné-Bissau. Os valores obtidos foram mais elevados para *env* comparativamente a *nef* do VIH-1, independentemente dos grupos de sequências analisados. Estas diferenças encontram-se de acordo com o estudo de van Harmelen *et al* (2001), no qual a região *env* possui distâncias genéticas superiores comparativamente a outras regiões genómicas, como *nef* ou *gag*, independentemente do subtipo viral. A variabilidade na região *env* pode ser fundamentada pela maior pressão seletiva exercida pelo sistema imunitário ou pela terapêutica antirretroviral

utilizada no tratamento (van Harmelen *et al*, 2001). Por outro lado, observou-se que independentemente da região genómica analisada (*env* ou *nef*) os valores das distâncias nucleotídicas foram superiores no grupo de sequências de referência G, o que era esperado pois a CRF02_AG foi introduzida à mais tempo na Guiné-Bissau, onde esta forma recombinante apresenta uma maior diversidade e evolução no tempo. Por sua vez, as sequências de referência G pertencem a indivíduos com a infeção VIH-1 estabelecida enquanto as sequências GP em estudo correspondem a mulheres jovens em que se prevê uma aquisição da infeção mais recentemente.

a) Sequências aminoacídicas de V3-V5

As sequências aminoacídicas da região V3-V5 revelaram um elevado grau de conservação nos resíduos e nos domínios estruturais considerados importantes para o desempenho das principais funções atribuídas à proteína Env do VIH-1.

Os resíduos de cisteína nas posições 1 e 35 de V3 responsáveis pela conformação da proteína encontraram-se conservados em todas as sequências GP analisadas.

Estudos publicados na literatura descrevem que a pressão seletiva atua no sentido de se manter um número mínimo de locais de N-glicosilação na proteína Env podendo ser adquiridos ao longo da infeção (Esteves *et al*, 2002). Com exceção da sequência GP211 onde foram identificados apenas 2 locais de N-glicosilação, verificou-se que para todas as sequências GP foram observados os 3 locais de N-glicosilação esperados.

Os recetores das quimiocinas CCR5 e CXCR4 desempenham papel relevante na patogenia da infeção por VIH-1, estando a sua utilização associada ao tempo de infeção no indivíduo. A utilização pela variante viral do co-recetor CCR5 é uma característica frequentemente observada numa fase inicial da infeção. A previsão de utilização dos co-recetores para as sequências GP em estudo revelou que a maioria (93%) utiliza o co-recetor CCR5, sugerindo que as infeções nestes indivíduos são relativamente recentes. De facto, neste estudo estão incluídas mulheres jovens (média 30,9 anos) infetadas por via heterossexual, o que pode corroborar a hipótese de não serem infeções VIH-1 estabelecidas. Apenas em dois casos (GP206 e GP236), foi prevista a utilização do co-recetor CXCR4, pelo que estas mulheres, comparativamente às outras, podem encontrar-se numa fase mais avançada da infeção

(Azevedo-Pereira & Costa-Santos, 2008). Na verdade, estas mulheres têm idade superior à média da população estudada, respectivamente 36 e 38 anos de idade.

b) Sequências aminoacídicas de Nef

A análise das sequências aminoacídicas de Nef revelou uma proteína potencialmente funcional não se encontrando truncada e apresentando a maioria dos motivos estruturais e funcionais conservados. A disrupção de alguns motivos ou resíduos pode ser compensada pela conservação de outros associados a funções semelhantes.

Observou-se que o sinal de meristilação (MGxxxS₁) encontra-se conservado em todas as sequências, pelo que a ancoragem de Nef na membrana citoplasmática não parece estar comprometida (Geyer *et al*, 1999). Também o motivo W₅₇ foi observado conservado na totalidade das sequências sugerindo que a capacidade de clivagem proteolítica da proteína Nef encontra-se preservada.

A maioria das alterações de aminoácidos observadas nas sequências GP foram substituições por resíduos com propriedades bioquímicas equivalentes (Anexo 2) e a sua elevada frequência nas sequências em estudo, sugere que possam ser polimorfismos associados a subtipos não-B. Por exemplo, o motivo M₂₀ e o motivo RRAE₂₁, ambos associados à regulação negativa da molécula MHC-I, não foram observados na forma que é descrita para estes motivos. Na posição 20 observou-se a substituição do resíduo de M por I em 74,1% das sequências analisadas, no entanto, ambos os aminoácidos possuem propriedades semelhantes (hidrófobas). O motivo RRAE₂₁ foi observado na forma R/KQTP em 85,2% das sequências analisadas. A substituição na posição 21 do resíduo R por K (ambos básicos e hidrófilos) e a substituição na posição 22 resíduo R por Q em 85% das sequências, sugere polimorfismos associados a subtipos não-B. De salientar que a esmagadora maioria das sequências GP foram classificadas em CRF02_AG.

No motivo ácido (EEEE₆₂) (Piguet *et al*, 1999) associado à regulação negativa do MHC-I, a maioria das alterações observadas correspondem à substituição por outros resíduos ácidos (E para D). Estas alterações foram descritas em estudos que sugerem que estas substituições

no motivo não comprometem as funções da proteína, admitindo que a conformação EEEE₆₂, é característica do subtipo B (Parreira *et al*, 2005). Na população em estudo não foi identificado nenhum caso de infecção pelo subtipo B.

O local de fosforilação de Nef pela cinase PKC (RPMTYK₇₇) (Kenelly & Krebs, 1991), apresentou-se conservado em 70,4% das sequências em estudo, no entanto, observaram-se substituições do resíduo Y₈₁ (aminoácido polar neutro) para S (igualmente polar neutro) na sequência GP244 e para F (aminoácido apolar) nas sequências GP208, GP209, GP219, GP221, GP235, GP246 e GP260 (n=7). Esta última substituição pode impedir a fosforilação de Nef por esta via e comprometer as respectivas funções da proteína.

Os motivos FDP₁₂₁ (Cohen *et al*, 2000), EE₁₅₄ (Benichou *et al*, 1994), E/DxxxLL₁₆₀ (Bresnahan *et al*, 1998) e DD₁₇₄ (Lu *et al*, 1998), associados no processo de ligação a moléculas mediadoras no tráfego celular apresentaram-se conservados não sendo observadas alterações que possam comprometer as funções da proteína na regulação negativa dos CD4 e MHC-I e II.

Os motivos de tetraprolina (PxxPxxPxRP₆₉) e prolina (PxxP₁₄₇) implicados na ligação aos domínios SH3 de diferentes cinases, apresentaram um elevado nível de conservação justificando a importância das vias de sinalização na infecção VIH-1. A única alteração observada foi na sequência GP235 com o motivo de prolina na forma PxxK₁₄₇, no entanto esta alteração não compromete a ligação de Nef com os domínios SH3, atendendo a que esta mesma sequência possui o motivo da tetrapolina totalmente conservado.

O motivo RR₁₀₅ envolvido na ligação à cinase PAK (Renkema *et al*, 1999), apresentou alterações em três sequências em estudo, no entanto, esta alteração foi por aminoácidos com propriedades bioquímicas semelhantes (R por K). O mesmo foi observado na análise do motivo envolvido na ligação à cinase C-Raf1 (DDPxxE₁₇₄) (Hodge *et al*, 1998), com alterações do resíduo na posição 174 de D para E e na posição 176 de P para A/D/E/K/S. A elevada frequência na substituição entre aminoácidos com propriedades ácidas (D/E) pode significar polimorfismos em sequências de subtipos não-B. Apenas uma única sequência (GP221) apresentou na posição 179 uma substituição de E por Q (resíduo polar neutro).

c) Sequências aminoacídicas da PR e RT e análise de mutações de resistência

As mutações de resistência são frequentemente detetadas a partir de plasma (ARN viral), sendo a detecção do ADN proviral apenas utilizado como alternativa, pois pode corresponder a material de arquivo e não representar a infecção do indivíduo no momento (Turriziani *et al*, 2010). Porém, um estudo realizado por Bon *et al*, (2007) revelou que as mutações que conferem resistência aos inibidores da transcriptase reversa e da protease podem ser detetadas com maior frequência em ADN extraído de CMSP do que em ARN viral extraído de plasma. Deste modo pode existir uma maior vantagem na utilização de ADN proviral extraído a partir de CMSP para a detecção de mutações de resistência uma vez que, o tratamento pode reduzir a carga viral para valores abaixo do limiar da técnica para detecção do ARN (Bon *et al*, 2007; Derache *et al*, 2015).

Nas sequências GP em estudo verificou-se uma elevada frequência de polimorfismos para a região PR, nomeadamente a mutação K20I e a mutação L89M, respetivamente em 92% e em 88% das sequências. A mutação L89M foi descrita como um polimorfismo natural de subtipos não-B. A mutação K20I é observada frequentemente em sequências do subtipo G e CRF02_AG (<https://hivdb.stanford.edu/dr-summary/comments/PI/>). Estes dados são corroborados por estudos de Abecasis *et al* (2006) e Santoro *et al* (2013), no qual estes polimorfismos naturais são predominantemente encontrados em subtipos não-B, com uma maior prevalência na CRF02_AG. Embora estas mutações sejam frequentes nas sequências GP em estudo não se encontram associadas à redução da suscetibilidade aos inibidores da protease, pelo que a sua identificação não confere uma resistência aos PI. No entanto, a combinação de várias mutações acessórias pode conduzir à redução da suscetibilidade aos PI. Curiosamente, a sequência GP256 que corresponde a uma mulher que não cumpriu terapia de prevenção e transmitiu o VIH-1 ao filho, apresentou as mutações K20I e L89M.

Em quatro (17%) sequências GP em estudo foram identificadas as mutações de resistência V11I, T74P e N88T. A mutação V11I observada nas sequências GP200 e GP216 foi descrita sendo selecionada em pacientes tratados com Darunavir (DRV). A mutação T74P foi identificada na sequência GP235 e representa uma mutação não polimórfica, encontrada principalmente em indivíduos tratados com mais que um inibidor da protease. A mutação N88T identificada na sequência GP205 é bastante rara, podendo ser associada à redução da

susceptibilidade do Nelfinavir (NFV) e Atazanavir (ATV) (<https://hivdb.stanford.edu/dr-summary/comments/PI/>).

Das sequências GP em estudo para a região RT foram identificadas cinco mutações: V90I, A98G, K103N, V106I e M184V. No entanto, a mutação V90I e a mutação V106I são consideradas mutações polimórficas com efeito reduzido na suscetibilidade aos inibidores da RT e foram encontradas respetivamente nas sequências GP208 e GP252. Em contraste, foram identificadas as mutações A98G e K103N, respetivamente nas sequências GP211 e GP262, que conferem elevado nível de resistência aos NNRTI (<https://hivdb.stanford.edu/dr-summary/resistance-notes/NRTI/>). O facto da mutação K103N ter sido detetada numa sequência GP em estudo, encontra-se de acordo com o estudo de Santoro *et al* (2013), que descreve uma maior prevalência desta mutação em indivíduos infetados por subtipos não-B, principalmente C, F e CRF02_AG. A mutação M184V foi a única mutação de resistência aos NRTI sendo identificada na sequência GP262. Esta mutação confere elevado grau de resistência ao 3TC e FTC e um grau médio de resistência ao ABC (<https://hivdb.stanford.edu/dr-summary/resistance-notes/NNRTI/>). De salientar que a sequência GP262 apresentou 3 mutações de resistência aos RTI (A98G, K103N e M184V) e não foram detetadas mutações de resistência nas sequências derivadas de mulheres que não cumpriram terapia antirretroviral de prevenção da transmissão mãe-filho do VIH-1.

De um modo geral, observou-se uma baixa frequência de mutações de resistência no grupo de mulheres estudadas. Foram identificados quatro (17%) casos de resistência aos PI e dois (8%) casos de resistência aos RTI. De facto, nos últimos anos observou-se uma evolução na eficácia dos regimes de terapia antiretroviral usados no tratamento e na prevenção da transmissão do VIH-1, que foram traduzidos por uma redução significativa do aparecimento de mutações de resistência.

6. Conclusão

A caracterização de diferentes regiões do genoma, concretamente a análise das regiões *env*, *nef*, PR e RT, permitiu classificar a infeção por VIH-1 no grupo de mulheres de nacionalidade Guineense a residir em Portugal, e consequentemente, permitiu comparar os resultados com a diversidade molecular encontrada no grupo de indivíduos nativos e residentes na Guiné-Bissau. Assim, com o presente estudo podemos concluir que:

- A maioria dos casos estudados são infeções CRF02_AG, sendo detetados apenas dois casos de infeção pelo sub-subtipo A1 (eventualmente puros) e ainda 3 casos de genomas mosaico, potencialmente URF (CRF02_AG/A/J). Deste modo, constatou-se que neste estudo existe uma diversidade genética semelhante à encontrada no grupo de indivíduos nativos e residentes na Guiné-Bissau. Estes resultados estão de acordo com o que se encontra descrito na epidemiologia de infeção na Guiné-Bissau. Por sua vez, a diversidade molecular existente na Guiné-Bissau é menor comparativamente a Portugal.
- Os casos de infeção identificados como potenciais formas recombinantes únicas necessitam da caracterização de um maior número de regiões genómicas, ou do genoma total, para obter uma classificação correta do vírus;
- Foram encontradas distâncias genéticas mais elevadas na região *env*, em comparação com a região *nef*, estando os dados de acordo com a literatura consultada. As distâncias genéticas do grupo de sequências de referência G (para as duas regiões genómicas estudadas), foram mais elevadas comparativamente com o grupo de sequências GP em estudo, uma vez que a CRF02_AG foi introduzida à mais tempo na Guiné-Bissau;
- A análise das sequências Env, Nef, PR e RT revelou elevada conservação dos principais motivos funcionais e estruturais das proteínas demonstrando a importância da sua preservação para a infecciosidade e replicação do vírus. Nas sequências analisadas foi detetada uma elevada prevalência de polimorfismos descritos para sequências de subtipos não-B;
- Observou-se a utilização do co-recetor CCR5 associado à infeção do VIH-1 na maioria dos casos, este facto sugere que as infeções dos indivíduos não estão estabelecidas. As mulheres incluídas neste estudo são jovens, pressupondo que a aquisição da infeção VIH-1 seja ser relativamente recente;

- Foi encontrada uma baixa prevalência de mutações de resistência aos inibidores da protease (V11I, T74P e N88T) e aos inibidores da transcriptase reversa (A98G, K103N e M184V), em mulheres que cumpriram regimes de terapia antirretroviral.

7. Referências Bibliográficas

- Abecasis AB, Deforche K, Bacheler LT, Mckenna P, Carvalho AP.** 2006. Investigation of baseline susceptibility to protease inhibitors in HIV-1 subtypes C , F , G and CRF02 _ AG. *Antivir Ther* **1**:581–589.
- Abecasis AB, Vandamme A, Lemey P.** 2009. Quantifying Differences in the Tempo of Human Immunodeficiency Virus Type 1 Subtype Evolution □. *J Virol* **83**:12917–12924.
- Abecasis AB, Wensing AMJ, Paraskevis D, Vercauteren J, Theys K, Vijver DAMC Van De, Albert J, Asjö B, Balotta C, Beshkov D, Camacho RJ.** 2013. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology* **10**:1–13.
- Adamson CS.** 2012. Protease-Mediated Maturation of HIV : Inhibitors of Protease and the Maturation Process. *Mol Biol Intern* 1-13
- Aiken C, Trono D.** 1995. Nef Stimulates Human Immunodeficiency Virus Type 1 Proviral DNA Synthesis. *J Virol* **69**:5048–5056.
- Andersson S, Norrgren H, Dias F, Biberfeld G, Albert J.** 1999. Molecular Characterization of Human Immunodeficiency Virus (HIV)-1 and -2 in Individuals from Guinea-Bissau with Single or Dual Infections: Predominance of a Distinct HIV-1 Subtype A/G Recombinant in West Africa. *Virology* **262**:312-320
- Antunes F, Doroana, M.** Infecção VIH-SIDA In Antunes, F., 2012. Manual sobre Doenças Infecciosas. Secção 7:203-250. Edição. ed. Permanyer Portugal.
- Arrildt K, Joseph S, Swanstrom R.** 2012. The HIV-1 Env Protein: A Coat of Many Colors. *Curr HIV/AIDS Rep* **9**:52–63.
- Arts EJ, Hazuda DJ.** 2012. HIV-1 Antiretroviral Drug Therapy. *Cold Spring Harb Perspect Med* **2**:1–23.
- Azevedo-Pereira J, Santos-Costa Q.** 2008. Os receptores das quimiocinas e a sua importância no ciclo replicativo do vírus da imunodeficiência humana- Implicações Clínicas e Terapêuticas. *Acta Med Port* **21**:497–504.
- Azevedo Pereira J.** Estrutura do HIV (Disponível em http://www.ff.ul.pt/~jazevedo/Site/Mest_CBF_files/HIV-JMAP.pdf - Acedido a 3 de Agosto de 2016)
- Bagaya BS, Vega JF, Tian M, Nickel GC, Li Y, Krebs KC, Arts EJ.** 2015. Functional bottlenecks for generation of HIV-1 intersubtype Env recombinants. *Retrovirology* **12**:1–17.
- Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vézinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220**:868–871.
- Bártolo I, Abecasis AB, Borrego P, Barroso H, Mccutchan F, Camacho R, Taveira N.** 2011. Origin and Epidemiological History of HIV-1 CRF14 _ BG. *PLoS One* **6**:1–9.
- Basmaciogullari S, Pizzato M.** 2014. The activity of Nef on HIV-1 infectivity. *Front Microbiol* **5**:1-12

- Benichou S, Bomsel M, Bodéus M, Durand H, Douté M, Letourneur F, Camonis J, Benarous R.** 1994. Physical interaction of the HIV-1 Nef protein with β -COP, a component of non-clathrin-coated vesicles essential for membrane traffic. *J Biol Chem.* **269**:30073–30076.
- Berger E a, Murphy PM, Farber JM.** 1999. Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease. *Annu Rev Immunol* **17**:657–700.
- Bresnahan PA, Yonemoto W, Ferrell S, Williams-herman D, Geleziunas R, Greene WC.** 1998. A dileucine motif in HIV-1 Nef acts as an internalization signal for CD4 downregulation and binds the AP-1 clathrin adaptor. *Curr Biol* **8**:1235–1238.
- Bon I, Alessandrini F, Borderi M, Gorini R, Re M.** 2007. Analysis of HIV-1 drug-resistant variants in plasma and peripheral blood mononuclear cells from untreated individuals: implications for clinical management. *New Microbiol* **30**:313-317.
- Butler IF, Pandrea I, Marx P a, Apetrei C.** 2007. HIV genetic diversity: biological and public health consequences. *Curr HIV Res* **5**:23–45.
- Carvalho A, Costa P, Triunfante V, Branca F, Rodrigues F, Santos CL, Correia-neves M, Saraiva M, Lecour H, Castro AG, Pedrosa J, Osório S.** 2015. Analysis of a Local HIV-1 Epidemic in Portugal Highlights Established Transmission of Non-B and Non-G Subtypes. *J Clin Microbiol* **53**:1506–1514.
- Castro-Nallar E, Pérez-Losada M, Burton GF, Crandall K a.** 2012. The evolution of HIV: Inferences using phylogenetics. *Mol Phylogenet Evol* **62**:777–792.
- Chen Y, Trono D, Camaur D.** 1998. The Proteolytic Cleavage of Human Immunodeficiency Virus Type 1 Nef Does Not Correlate with Its Ability To Stimulate Virion Infectivity. *J Virol* **72**:3178–3184.
- Chen R, Yokoyama M, Sato H, Reilly C, Mansky LM.** 2005. Human Immunodeficiency Virus Mutagenesis during Antiviral Therapy : Impact of Drug-Resistant Reverse Transcriptase and Nucleoside and Nonnucleoside Reverse Transcriptase Inhibitors on Human Immunodeficiency Virus Type 1 Mutation Frequencies **79**:12045–12057.
- Cho MW, Lee MK, Carney MC, Berson JF, Doms RW, Martin MA.** 1998. Identification of Determinants on a Dualtropic Human Immunodeficiency Virus Type 1 Envelope Glycoprotein That Confer Usage of CXCR4. *J Virol* **72**:2509–2515.
- Clapham, P; McKnight Á.** 2002. Cell surface receptors , virus entry and tropism of primate lentiviruses. *J Gen Virol* **83**:1809–1829.
- Clavel F, Guetard D, Brunvezinet F, Chamaret S, Ray M, Santos-Ferreira M, Laurent A, Dauguet C, Katlama C, Rouzioux C.** 1986. Isolation of a New Human Retrovirus from West-African Patients with AIDS. *Science* **233**:343– 346.
- Cohen GB, Rangan VS, Chen BK, Smith S, Baltimore D.** 2000. The Human Thioesterase II Protein Binds to a Site on HIV-1 Nef Critical for CD4 Down-regulation *. *J Biol Chem* **275**:23097–23105.
- Cohen MS, Hellmann N, Levy JA, Decock K, Lange J, Hill C, Hill C, Carolina N, Foundation MG.** 2008. Review series The spread , treatment , and prevention of HIV-1 : evolution of a global pandemic. *J Clin Invest* **118**:1244–1254.

- Cormier EG, Dragic T.** 2002. The Crown and Stem of the V3 Loop Play Distinct Roles in Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Interactions with the CCR5 Coreceptor. *J Virol* **76**:8953–8957.
- Cornett JK, Kirn TJ.** 2013. Laboratory Diagnosis of HIV in Adults : A Review of Current Methods. *Clin Infect Dis* 1–7.
- Delwart L, Pan H, Sheppard W, Wolpert D, Neumann A, Korber B, Mullins J.** 1997. Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. *J. Virol.* **10**:7498-7508.
- Departamento de Doenças Infecciosas do INSA, Programa Nacional para a Infecção VIH/SIDA.** 2015. Infecção VIH / SIDA : a situação em Portugal a 31 de dezembro de 2014. Relatório 2014, Instituto Nacional de Saúde Ricardo Jorge. **146**, 42-45
- Derache A, Shin H, Balamane M, White E, Israelski D, Klausner J, Freeman A, Katzenstein D.** 2015. HIV Drug Resistance Mutations in Proviral DNA from a Community Treatment Program. *PLoS One* **10**:e0117430
- Eberle J, Gürtler L.** 2012. Basic Principles of HIV HIV Types , Groups , Subtypes and Recombinant Forms : Errors in Replication , Selection Pressure and Quasispecies. *Intervirology* **55**:79–83.
- Engelman A, Cherepanov P.** 2012. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat Rev Microbiol* **10**:279–290.
- Esbjörnsson J, Mild M, Månsson F, Norrgren H, Medstrand P.** 2011. HIV-1 molecular epidemiology in Guinea-Bissau, West Africa: Origin, demography and migrations. *PLoS One* **6**:e17025.
- Esteves A, Parreira R, Venenno T, Franco M, Piedade J, Germano De Sousa J, Canas-Ferreira WF.** 2002. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Res Hum Retroviruses* **18**:313–325.
- Fearon M.** 2005. The laboratory diagnosis of HIV infections. *J Infect Dis Med Microbiol* **16**:26–30.
- Flynn WF, Chang MW, Tan Z, Oliveira G, Yuan J, Okulicz F, Torbett BE, Levy RM.** 2015. Deep Sequencing of Protease Inhibitor Resistant HIV Patient Isolates Reveals Patterns of Correlated Mutations in Gag and Protease. *PLOS Comput Biol* 1–27.
- Foda M, Harada S, Maeda Y.** 2001. Role of V3 Independent Domains on a Dualtropic Human Immunodeficiency Virus Type 1 (HIV-1) Envelope gp120 in CCR5 Coreceptor Utilization and Viral Infectivity. *Microbiol Immunol* **45**:521–530.
- Foster G, Ambrose J, Hué S, Delpéch VC, Fearnhill E, Foster GM, Ambrose JC.** 2014. Novel HIV-1 Recombinants Spreading across Multiple Risk Groups in the United Kingdom : The Identification and Phylogeography of Circulating Recombinant Form. *PLoS One* **9**:1–10.
- Freed EO.** 2001. HIV-1 Replication. *Somat Cell Mol Genet* **26**:13–33.
- Freitas FB, Esteves A, Parreira R.** 2013. Novel Multiregion Hybridization Assay for the Identification of the Most Prevalent Genetic Forms of the Human Immunodeficiency Virus Type 1 Circulating in Portugal. *AIDS Res Hum Retroviruses* **29**:318–328.

- Frieden TR, Foti KE, Mermin J.** 2015. Medicine and Society Applying Public Health Principles to the HIV Epidemic — How Are We Doing ? *N Engl J Med* **373**:2281–2287.
- Friedrich BM, Dziuba N, Li G, Endsley M a., Murray JL, Ferguson MR.** 2011. Host factors mediating HIV-1 replication. *Virus Res* **161**:101–114.
- Fun A, Maarseveen NM Van, Pokorná J, Maas REM, Schipper PJ, Konvalinka J, Nijhuis M.** 2011. HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. *Retrovirology* **8**:70.
- Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH.** 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes*. *Nature* **397**:436–441.
- Geyer M, Fackler OT, Peterlin BM.** 2001. Structure – function relationships in HIV-1 Nef. *EMBO Rep* **2**:580–585.
- Greenway A, Holloway G, McPhee D, Ellis P, Cornall A, Lidman M.** 2003. HIV-1 Nef control of cell signalling molecules : multiple strategies to promote virus replication. *J Biosci* **28**:323–335.
- Gupta RM, Sahni AK, Jena J, Nema SK.** 2005. Genomic Diversity of Human Immunodeficiency Viruses. *Med J Armed Forces India* **61**:267–270.
- Gutiérrez-Rivas M, Menéndez-Arias L.** 2001. A mutation in the primer grip region of HIV-1 reverse transcriptase that confers reduced fidelity of DNA synthesis. *Nucleic Acids Res* **29**:4963–4972.
- Hahn BH, Shaw GM, De Cock KM, Sharp PM.** 2000. AIDS as a zoonosis: scientific and public health implications. *Science* **287**:607–614.
- Hemelaar J.** 2012. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med* **18**:182–192.
- Hemelaar J, Gouws E, Ghys PD.** 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* **20**:13–23.
- HIV Drug Resistance Database.** 2016. NNRTI Resistance Mutation Comments. (Disponível em <https://hivdb.stanford.edu/dr-summary/resistance-notes/NNRTI/> - Acedido a 10 de Setembro de 2016)
- HIV Drug Resistance Database.** 2016. NRTI Resistance Mutation Comments. (Disponível em <https://hivdb.stanford.edu/dr-summary/resistance-notes/NRTI/> - Acedido a 10 de Setembro de 2016)
- HIV Drug Resistance Database.** 2016. PI Resistance Mutation Comments. (Disponível em <https://hivdb.stanford.edu/dr-summary/comments/PI/> - Acedido a 10 de Setembro de 2016)
- HIV sequence database.** 2015. Distribution of all HIV-1 sequences: Guinea-Bissau. (Disponível em http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp?organism=HIV-1®ion=®ion=gw&form=all&prev_region=pt&path=%2Fhome%2Fhttpd%2Fhtdocs%2Fcomponents%2Fsequence%2FHIV%2Fgeo- Acedido a 7 de Setembro de 2016)
- HIV sequence database.** 2015. Distribution of all HIV-1 sequences: Portugal. (Disponível em

- http://www.hiv.lanl.gov/components/sequence/HIV/geo/geo.comp?organism=HIV-1®ion=world®ion=pt&form=all&prev_region=world&path=%2Fhome%2Fhttpd%2Fhtdocs%2Fcomponents%2Fsequence%2FHIV%2Fgeo - Acedido a 5 de Setembro de 2016)
- HIV sequence database.** 2015. HIV Circulating Recombinant Forms (CRFs). (Disponível em <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html> - Acedido a 11 de Maio de 2016)
- HIV sequence database.** 2016. N-GlycoSite. (Disponível em <http://www.hiv.lanl.gov/content/sequence/GLYCOSITE/glycosite.html> - 27 de Agosto de 2016)
- Hodge DR, Dunn KJ, Pei GK, Chakrabarty MK, Heidecker G, Lautenberger JA, Samuel KP.** 1998. Binding of c-Raf1 Kinase to a Conserved Acidic Sequence within the Carboxyl-terminal Region of the HIV-1 Nef Protein *. *J Biol Chem* **273**:15727–15733.
- Hornak V, Simmerling C, Brook S, Brook S.** 2007. Targeting structural flexibility in HIV-1 protease inhibitor binding. *Drug Discov Today* **12**:132–138.
- Hou T, Mclaughlin W, Wang W.** 2008. resistance. *Proteins* **71**:1163–1174.
- International Committee on Taxonomy of Viruses.** 2016. Virus Taxonomy. (Disponível em http://www.ictvonline.org/virusTaxonomy.asp?taxnode_id=20040149 - Acedido a 8 de Maio de 2016)
- Jensen MA, Li F, Wout B Van, Nickle DC, Shriner D, He H, Mclaughlin S, Shankarappa R, Margolick JB, Mullins JL.** 2003. Improved Coreceptor Usage Prediction and Genotypic Monitoring of R5-to-X4 Transition by Motif Analysis of Human Immunodeficiency. *J Virol* **77**:13376–13388.
- Jespersen S, Tolstrup M, Hønge B, Medina C, Té D, Ellermann-Eriksen S, Ostergaard L, Wejse C, Laursen A.** 2015. High level of HIV-1 drug resistance among patients with HIV-1 and HIV-1/2 dual infections in Guinea-Bissau. *Virol J* **12**:41–47.
- Kandathil A, Ramalingam S, Kannangai R, David S, Sridharan G.** 2005. Molecular epidemiology of HIV infection. *Indian J Med Res* **121**:333–344.
- Kennelly PJ, Krebs EG.** 1991. Consensus Sequences as Substrate Specificity Determinants for Protein Kinases and Protein Phosphatases. *J Biol Chem* **266**:15555–15556.
- Khayat R, Lee J, Julien J-P, Cupo A, Klasse PJ, Sanders R, Moore J, Wilson I, Ward A.** 2013. Structural Characterization of Cleaved, Soluble HIV-1 Envelope Glycoprotein Trimers. *J Virol* **87**:9865–9872.
- Kimura M.** 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **2**:111-20
- Kirchhoff F, Schindler M, Specht A, Arhel N, Münch J.** 2008. Review Role of Nef in primate lentiviral immunopathogenesis. *Cell Mol Life Sci* **65**:2621–2636.
- Krogstad P.** 2003. Molecular Biology of the Human Immunodeficiency Virus: Current and Future Targets for Intervention. *Semin Pediatr Infect Dis* **14**:258–268.
- Kwong PD, Wyatt R, Majeed S, Robinson J, Sweet RW, Sodroski J, Hendrickson WA.** 2000.

Structures of HIV-1 gp120 Envelope Glycoproteins from Laboratory-Adapted and Primary Isolates. *Structure* **8**:1329–1339.

Larsen O, Silva Z, Sandström A, Andersen PK, Andersson S, Poulsen A, Melbye M, Dias F, Naclér A, Aaby P. 1998. Declining HIV-2 prevalence and incidence among men in a community study from Guinea-Bissau. *AIDS* **12**:1707–1714.

Levy J. 2009. HIV pathogenesis: 25 years of progress and persistent challenges. *AIDS* **23**:147–160.

Li P, Wang T, Buckley KA, Popov S, Ruprecht RM. 2005. Phosphorylation of HIV Nef by cAMP-dependent protein kinase. *Virology* **331**:367–374.

Lu X, Yu H, Liu S, Brodsky FM, Peterlin BM. 1998. Interactions between HIV1 Nef and Vacuolar ATPase Facilitate the Internalization of CD4. *Immunity* **8**:647–656.

Lv Z, Chu Y, Wang Y. 2015. HIV protease inhibitors: a review of molecular selectivity and toxicity. *HIV/AIDS – Res Palliative Care* **7**:95–104.

Maartens G, Celum C, Lewin SR. 2014. HIV infection: Epidemiology, pathogenesis, treatment, and prevention. *Lancet* **384**:258–271.

Malim MH, Emerman M. 2008. HIV-1 Accessory Proteins - Ensuring Viral Survival in a Hostile Environment. *Cell Host Microb* **3**:388–398.

Matos PM, Andreu D, Santos NC, Gutie R. 2014. Structural requirements of glycosaminoglycans for their interaction with HIV-1 envelope glycoprotein gp120. *Arch Virol* **159**:555–560.

Max Planck Institut Informatik. 2000. Geno2Pheno: Resistance. (Disponível em <http://www.geno2pheno.org/> - Acedido a 1 de Setembro de 2016)

McCutchan F, Sankale JL, M'Boup S, Kim B, Tovanabutra S, Hamel D, Brodine S, Kanki P, Birx D. 2004. *AIDS Res Hum Retroviruses*. **20**: 819-826

Menéndez-Arias L. 2010. Molecular basis of human immunodeficiency virus drug resistance: An update. *Antiviral Res* **85**:210–231.

Mulky A, Sarafianos SG, Jia Y, Arnold E, Kappes JC. 2005. Identification of Amino Acid Residues in the Human Immunodeficiency Virus Type-1 Reverse Transcriptase Tryptophan-repeat Motif that are Required for Subunit Interaction Using Infectious Virions. *J Mol Biol* **349**:673–684.

Mulky A, Vu BC, Conway JA, Hughes SH, Kappes JC. 2007. Analysis of Amino Acids in the β 7 – β 8 Loop of Human Immunodeficiency Virus Type 1 Reverse Transcriptase for their Role in Virus Replication. *J Mol Biol* **365**:1368–1378.

Mullins Computational Biology Tools. 2009. Web PSSM. (Disponível em <https://indra.mullins.microbiol.washington.edu/webpssm/> - Acedido a 24 de Agosto de 2016

Niculescu I, Paraschiv S, Paraskevis D, Abagiu A. 2015. Recent HIV-1 Outbreak Among Intravenous. *AIDS Res Hum Retroviruses* **31**:488–495.

Njai H, Gali Y, Vanham G, Clybergh C, Jennes W, Vidal N, Butel C, Mpoudi-Ngolle E, Peeters M, Ariën K. 2006. The predominance of Human Immunodeficiency Virus type 1 (HIV-1) circulating recombinant form 02 (CRF02_AG) in West Central Africa may be related to its

replicative fitness. *Retrovirology* **3**:40

Osmanov S, Pattou C, Walker N, Schwardlander B, Esparza J, Characterization the W-UN for HI and. 2002. Estimated Global Distribution and Regional Spread of HIV-1 Genetic Subtypes in the Year 2000. *J Acquir Immune Defic Syndr* **29**:184–190.

Palm AA, Esbjörnsson J, Månsson F, Kvist A, Isberg P, Biague A, José Z, Jansson M, Norrgren H, Medstrand P. 2013. Faster Progression to AIDS and AIDS-Related Death Among Seroincident Individuals Infected With Recombinant HIV-1 A3 / CRF02_AG Compared With Sub-subtype A3. *J Infect Dis* 1–8.

Palm AA, Esbjörnsson J, Månsson F, Biague A, Silva Z, Norrgren H, Jansson M, Medstrand P. 2015. Co-circulation of several similar but unique HIV-1 recombinant forms in Guinea-Bissau revealed by near full-length genomic sequencing. *AIDS Res Hum Retroviruses* 1-32

Palma A, Araújo F, Duque V, Borges F, Paixa MT. 2007. Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect Genet Evol* **7**:391–398.

Parreira R, Pádua E, Piedade J, Venenno T, Paixão MT, Esteves A. 2005. Genetic Analysis of Human Immunodeficiency Virus Type 1 nef in Portugal : Subtyping , Identification of Mosaic Genes , and Amino Acid Sequence Variability. *J Med Virol* **77**:8–16.

Pereyra F, Palmer S, Miura T, Block BL, Wiegand A, Rothchild AC, Baker B, Rosenberg R, Cutrell E, Seaman MS, Coffin JM, Walker BD. 2009. Persistent Low-Level Viremia in HIV-1 Elite Controllers and Relationship to Immunologic Parameters. *J Infect Dis* **200**:984–990.

Perrin L, Kaiser L, Yerly S. 2003. Travel and the spread of HIV-1 genetic variants. *Lancet* **3**:22–27.

Piguet V, Gu F, Foti M, Demareux N, Gruenberg J, Carpentier J, Trono D. 1999. Nef-Induced CD4 Degradation : A Diacidic-Based Motif in Nef Functions as a Lysosomal Targeting Signal through the Binding of β -COP in Endosomes. *Cell Press* **97**:63–73.

Pinto M, Struchiner C. 2006. HIV-1 diversity : a tool for studying the pandemic. *Cad Saúde Pública* **22**:473–484.

Plantier J, Leoz M, Dickerson JE, Damond F, Robertson DL. 2016. A new human immunodeficiency virus derived from gorillas. *Nat Med* **15**:871–872.

Pokorná J, Machala L, Řezáčová P, Konvalinka J. 2009. Current and Novel Inhibitors of HIV Protease. *Viruses* **1**:209-1239

Renkema GH, Manninen A, Mann DA, Harris M, Saksela K. 1999. Identification of the Nef-associated kinase as p21-activated kinase 2. *Curr Biol* **9**:1407–1411.

Requejo HIZ. 2006. Worldwide molecular epidemiology of HIV. *Rev Saude Publica* **40**:331–345.

Resch W, Hoffman N, Swanstrom R. 2001. Improved Success of Phenotype Prediction of the Human Immunodeficiency Virus Type 1 from Envelope Variable Loop 3 Sequence Using Neural Networks. *Virology* **288**:51–62.

Saitou N, Nei M. 1987. The Neighbor-joining Method : A New Method for Reconstructing

Phylogenetic Trees. *Mol Biol Evol* **4**:406–425.

Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* **74**:5463–5467.

Santoro MM, Perno CF. 2013. HIV-1 Genetic Variability and Clinical Implications. *ISRN Microbiol* **2013**:1–20.

Sarafianos SG, Marchand B, Das K, Himmel D, Michael A, Hughes SH, Arnold E. 2009. Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition. *J Mol Biol* **385**:693–713.

Shafer RW, Rhee S, Pillay D, Miller V, Sandstrom P, Schapiro JM, Kuritzkes DR, Bennett D. 2007. HIV-1 protease and reverse transcriptase mutations for drug resistance surveillance. *AIDS* **21**:215–223

Sharma P, Nurpeisov V, Schinazi R. 2005. Retrovirus reverse transcriptases containing a modified YXDD motif. *Antivir Chem Chemother* **16**:169–182.

Sharp PM, Hahn BH. 2001. Origins of HIV and the AIDS epidemic. *Cold Spring Harb Perspectives Med* **1**:a006841.

Sierra S, Kupfer B, Kaiser R. 2005. Basics of the virology of HIV-1 and its replication. *J Clin Virol* **34**:233–244.

Silva Z, Oliveira I, Andersen A, Dias F, Rodrigues A, Holmgren B, Andersson S, Aaby P. 2008. Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing? *AIDS* **22**:1195–1202

Simon V, Ho D, Karim Q. 2006. HIV/AIDS epidemiology, pathogenesis, prevention and treatment. *Lancet* **368**:489–504.

Simonsen JN, Fowke KR, MacDonald KS, Plummer FA. 1998. HIV pathogenesis : mechanisms of susceptibility and disease progression. *Curr Biol* **1**:423–429.

Singh P, Kaur G, Sharma G, Mehra NK. 2008. Immunogenetic basis of HIV-1 infection, transmission and disease progression. *Vaccine* **26**:2966–2980.

Sluis-Cremer N, Kempner E, Parniak M. 2003. Structure – activity relationships in HIV-1 reverse transcriptase revealed by radiation target analysis. *Protein Sci* **12**:2081–2086.

Sluis-Cremer N, Arion D, Abram M, Parniak M. 2004. Proteolytic processing of an HIV-1 pol polyprotein precursor : insights into the mechanism of reverse transcriptase p66 / p51 heterodimer formation. *Int J Biochem Cell Biol* **36**:1836–1847.

Sluis-Cremer N, Tachedjian G. 2008. Mechanisms of inhibition of HIV replication by nonnucleoside reverse transcriptase inhibitors. *Virus Res.* **134**: 147–156

Smyth RP, Davenport MP, Mak J. 2012. The origin of genetic diversity in HIV-1. *Virus Res* **169**:415–429.

Stebbing J, Moyle G. 2003. The Clades of HIV : Their Origins and Clinical Significance. *AIDS Rev* **5**:205–213.

Steege K, Demecheleer E, Cabooter N De, Temmerman M, Ndumbe P, Mandaliya K, Plum

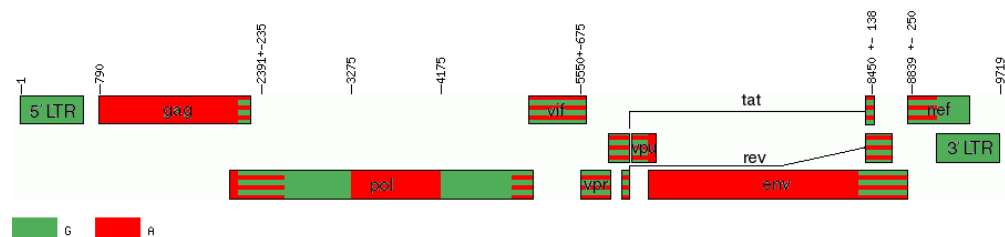
- J, Verhofstede C.** 2006. A sensitive in-house RT-PCR genotyping system for combined detection of plasma HIV-1 and assessment of drug resistance. *J Virol Methods* **133**:137–145.
- Tachedjian G, Aronson HG, Santos MDL, Seehra J, McCoy JM, Goff SP.** 2003. Role of Residues in the Tryptophan Repeat Motif for HIV-1 Reverse Transcriptase Dimerization. *J Mol Biol* **2836**:381–396.
- Taylor, Barbara; Sobieszczyk, Magdalena; McCutchan, Francine; Hammer S.** 2008. The Challenge of HIV-1 Subtype Diversity. *N Eng J Med* **358**:1590–1602.
- Tebit DM, Arts EJ.** 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. *Lancet Infect Dis* **11**:45–56.
- The Aaron Diamond AIDS Research Center.** 2012. The HIV Replication Cycle http://www.adarc.org/hiv_replication_632.html - Acedido a 3 de Maio de 2016)
- Thomson MM, Pérez-álvarez L, Nájera R.** 2002. Reviews Molecular epidemiology of HIV-1 genetic forms therapy. *Lancet* **2**:461–471.
- Todd S, Anderson C, Jolly DJ, Craik CS.** 2000. HIV protease as a target for retrovirus vector-mediated gene therapy. *Biochim Biophys Acta* **1477**:168–188.
- Tomasselli AG, Heinrikson RL.** 2000. Targeting the HIV-protease in AIDS therapy : a current clinical. *Biochim Biophys Acta* **1477**:189–214.
- Turner B, Summers M.** 1999. Strutural biology of HIV. *J Mol Biol* **285**: 1-32
- Turriziani O, Andreoni M, Antonelli G.** 2010. Resistant viral variants in cellular reservoirs of human immunodeficiency virus infection. *Clin Microbiol Infect* **16**:1518–1524
- van Harmelen J, Williamson C, Kim B, Morris L, Carr J, Karim SS, McCutchan F.** 2001. Characterization of full-length HIV type 1 subtype C sequences from South Africa. *AIDS Res Hum Retroviruses*. **17**:1527-1531.
- Vaz Sena F.** 2010. Prevalência e caracterização genética do Vírus da Imunodeficiência Humana em indivíduos atendidos no Hospital de Cumura em Bissau. Dissertação de Mestrado em Microbiologia Aplicada. Faculdade de Ciências da Universidade de Lisboa.
- Verveka V, Bauerova H, Zábranský A, Lang J, Ruml T, Pichova I, Hrabal R.** 2003. Three-dimensional Structure of a Monomeric Form of a Retroviral Protease. *J Mol Biol* **333**:771–780.
- Videira e Castro S.** 2010. Anti-retrovirais inibidores enzimáticos de vírus da imunodeficiência humana tipo 1 (HIV-1). Dissertação de Mestrado em Microbiologia Médica. Universidade Nova de Lisboa.
- Xiao L, Owen SM, Goldman I, Lal AA, Jacques J, Goudsmit J, Lal RB.** 1998. CCR5 Coreceptor Usage of Non-Syncytium-Inducing Primary HIV-1 Is Independent of Phylogenetically Distinct Global HIV-1 Isolates : Delineation of Consensus Motif in the V3 Domain That Predicts CCR-5 Usage. *Virology* **240**:83–92.
- World Health Organization.** 2016. Global Health Observatory (GHO) data (Disponível em <http://www.who.int/gho/hiv/en/> - Acedido a 30 de Março de 2016)

8. Anexos

Anexo 1. Formas recombinantes em circulação relevantes para este estudo**CRF02_AG**

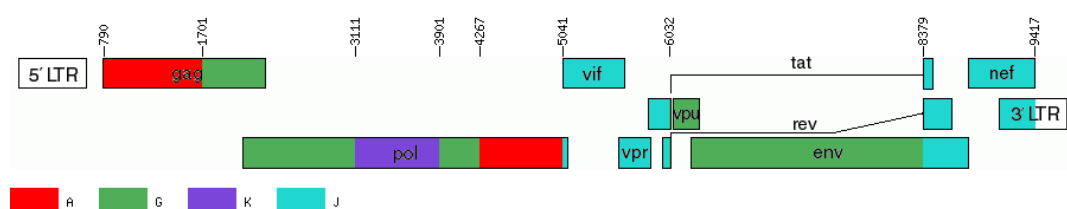
Subtipos: A e G

Pontos de recombinação: [1 a 789-G]; [790 a 2155-A]; [2156 a 2625-A/G]; [2626 a 3274-G]; [3275 a 4174-A]; [4175 a 4874-G]; [4875 a 6224-A/G]; [6225 a 8311-A]; [8312 a 8588-A/G]; [8589 a 9089 A/G]; [9090 a 9719-G]

**CRF06_cpx**

Subtipos: A, G, K e J

Pontos de recombinação: [790 a 1700-A]; [1701 a 3110-3]; [3111 a 3900-K]; [3901 a 4266-G]; [4267 a 5040-A]; [5041 a 6031-J]; [6032 a 8378-G]; [8378 a 9417-J].



Anexo 2: Classificação dos aminoácidos

Classificação dos aminoácidos		
Abreviatura do aminoácido	Nome do aminoácido	Propriedades Bioquímicas
A	Alanina	Apolar Hidrófobo
F	Fenilalanina	
I	Isoleucina	
L	Leucina	
M	Metionina	
P	Prolina	
V	Valina	
W	Triptofano	Neutro Hidrófilo
C	Cisteína	
G	Glicina	
N	Asparagina	
S	Serina	
Q	Glutamina	
T	Treonina	
Y	Tirosina	Ácido Hidrófobo
D	Ácido aspártico	
E	Ácido glutâmico	Básico Hidrófilo
H	Histidina	
K	Lisina	
R	Arginina	